12-2021

Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations

Sedeeq Al-khazraji
sha6709@rit.edu

Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations

by

Sedeeq Al-khazraji

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Computing and Information Sciences

B. Thomas Golisano College of Computing and
Information Sciences

Rochester Institute of Technology
Rochester, New York
December, 2021

Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations

by

Sedeeq Al-khazraji

Committee Approval:

We, the undersigned committee members, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Computing and Information Sciences.

Dr. Matt Huenerfauth Dissertation Advisor	Date
Dr. Cecilia Alm Dissertation Committee Member	Date
Dr. Kristen Shinohara Dissertation Committee Member	Date
Dr. Raja Kushalnagar Dissertation Committee Member	Date
Dr. Jai Kang Dissertation Defense Chairperson	Date

Certified by:

Dr. Pengcheng Shi Ph.D. Program Director, Computing and Information Sciences	Date

Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations

by

Sedeeq Al-khazraji

Submitted to the
B. Thomas Golisano College of Computing and Information Sciences Ph.D.
Program in Computing and Information Sciences
in partial fulfillment of the requirements for the
Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

While there are many Deaf or Hard of Hearing (DHH) individuals with excellent reading literacy, there are also some DHH individuals who have lower English literacy. American Sign Language (ASL) is not simply a method of representing English sentences. It is possible for an individual to be fluent in ASL, while having limited fluency in English. To overcome this barrier, we aim to make it easier to generate ASL animations for websites, through the use of motion-capture data recorded from human signers to build different predictive models for ASL animations; our goal is to automate this aspect of animation synthesis to create realistic animations.

This dissertation consists of several parts: **PART I**, defines key terminology for timing and speed parameters, and surveys literature on prior linguistic and computational research on ASL. Next, the motion-capture data that our lab recorded from human signers is discussed, and details are provided about how we enhanced this corpus to make it useful for speed and timing research. Finally, we present the process of adding layers of linguistic annotation and processing this data for speed and timing research.

PART II presents our research on data-driven predictive models for various speed and timing parameters of ASL animations. The focus is on predicting the (1) existence of pauses after each ASL sign, (2) predicting the time duration of these pauses, and (3) predicting the change of speed for each ASL sign within a sentence. We measure the quality of the proposed models by comparing our

models with state-of-the-art rule-based models. Furthermore, using these models, we synthesized ASL animation stimuli and conducted a user-based evaluation with DHH individuals to measure the usability of the resulting animation.

Finally, **PART III** presents research on whether the timing parameters individuals prefer for animation may differ from those in recordings of human signers. Furthermore, it also includes research to investigate the distribution of acceleration curves in recordings of human signers and whether utilizing a similar set of curves in ASL animations leads to measurable improvements in DHH users' perception of animation quality.

Acknowledgments

I would like to express my sincere gratitude to my advisors Dr. Matt Huenerfauth for his endless guidance, encouragement, and patience during my doctoral studies. It was a great pleasure to work under his supervision and learn from him. His guidance has helped me not only in the time of research and writing of this thesis but also at times when I doubted myself and my intuitions. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I could not have imagined having a better advisor other than Dr. Matt Huenerfauth.

I would also like to thank my dissertation committee members, Drs. Cissi Ovesdotter Alm, Kristen Shinohara, and Raja Kushalnagar for serving on my committee as well as their encouragement and insightful feedback.

Last but not least, I would like to thank my parents, Hassan Albana Ali and Latifa Saleh, who have always loved me unconditionally. I would like to thank my family, especially my aunts (Refaa and Taymea), my brothers, and my friends, for believing in me, always.

To my wife Razan

and my daughters Talia and Maria

My love, my life, my light.

Contents

List of Figures xiii
List of Tables xviii

1 Introduction **1**

1.1 Motivation for Research on Synthesizing ASL Animation 1
1.2 Focus of This Dissertation 3
1.3 Contributions of This Dissertation 5
1.4 Structure of This Dissertation 8

PART I: UNDERSTANDING THE GENERATION OF ASL ANIMATIONS AND BUILDING DATASET **9**

PROLOGUE TO PART I **10**

2 Key Concepts for ASL Speed and Timing **11**

2.1 Sequential Representations of ASL Signs and Animation 12
2.2 General Pipeline for Sign Language Generation 14
2.3 Speed, Acceleration, and Biomechanics of Hand Movement 17
2.4 Speed and Timing in ASL 18
2.4.1 Fundamental Rate (Words Per Second, Overall) 20
2.4.2 Base Duration (of a Word in a Lexicon) 20
2.4.3 Differential Signing Rate (of a Word as Performed) 21
2.5 Pauses in ASL 21

2.5.1 Pause Insertion . . . 22
2.5.2 Pause Duration . . . 22
2.6 Typical Speed and Timing Values for ASL and for Spoken English . . . 23
2.7 Why Speed and Timing Parameters are Important for Animators . . . 23
2.8 Summary . . . 25

3 Prior Methods of ASL Generation and Synthesis 26
3.1 Related Work in Speech Synthesis . . . 27
3.2 Data-Driven Sign Language Animation . . . 28
3.3 Speed and Timing in Sign Language Animations . . . 31
3.3.1 Fundamental Rate in Animation Systems . . . 34
3.3.2 Base Duration in Animation Systems . . . 34
3.3.3 Differential Signing Rate in Animation Systems . . . 35
3.3.4 Pause Insertion in Animation Systems . . . 35
3.3.5 Pause Duration in Animation Systems . . . 36
3.4 Evaluating the Quality of Output ASL . . . 36
3.4.1 Dataset-Based Validation Evaluation . . . 37
3.4.2 User-Based Evaluation of Animations . . . 38
3.5 Conclusion . . . 39

4 Creating the ASL Speed and Timing Dataset 40
4.1 Original Motion-Capture Corpus . . . 41
4.1.1 What was Good About This Existing Corpus for ASL Speed and Timing Research? . . . 45
4.1.2 What Limitations did This Existing Corpus have for ASL Speed and Timing Research? . . . 45
4.2 Adding Additional Annotation and Building an ELAN Motion Capture Dataset . . . 46
4.2.1 Dataset Annotation . . . 49
4.3 Data Extracting and Pre-processing . . . 50
4.4 Conclusion . . . 55

EPILOGUE FOR PART I 57

PART II: MODELING AND SYNTHESIS OF ASL ANIMATION 59

PROLOGUE TO PART II 60

5 Selecting Data-Driven Models of ASL Speed & Timing 62

5.1 Feature Engineering . . . 65
5.2 Models Overview . . . 67
5.3 Important Assumptions Before Modeling . . . 67
5.3.1 Assumptions Used to Estimate Differential Speed . . . 67
5.3.2 Assumptions Used to Estimate Pause Insertion & Duration . . . 68
5.4 Pause Insertion Modeling . . . 69
5.4.1 Design of Pause Insertion Model . . . 69
5.4.2 Features used for Pause Insertion Model . . . 70
5.4.3 Pause Insertion Cross-Validation Model Evaluation . . . 71
5.5 Differential Rate Modeling . . . 72
5.5.1 Design of Differential Rate Model . . . 73
5.5.2 Feature Used for Differential Rate Model . . . 73
5.5.3 Differential Rate Cross-Validation Model Evaluation . . . 73
5.6 Pause Duration Modeling . . . 74
5.6.1 Design of Pause Duration Model . . . 74
5.6.2 Feature Engineering for Pause Duration modeling . . . 74
5.6.3 Pause Duration Cross-Validation Model . . . 75
5.7 Dataset-Based Comparison to State-of-the-Art Model . . . 75
5.8 Conclusion . . . 78

6 Model Evaluation 80

6.1 User-Study with Subjective Feedback . . . 81
6.2 Participants . . . 81
6.3 Procedure and Data Collection . . . 82
6.4 Feedback of the User Study . . . 83
6.5 Conclusion . . . 87

EPILOGUE FOR PART II 88

PART III: USER PREFERENCES FOR SPEED, TIMING, AND ACCELERATION IN ASL 89

PROLOGUE TO PART III 90

7 Empirical Investigation ... 92

7.1 Prior Work Relevant to This Chapter . . . 94
7.2 Background . . . 95
7.3 Pilot Study . . . 98
7.3.1 Method . . . 98
7.4 Initial Five-Way Comparison Study . . . 108
7.4.1 Method . . . 110
7.5 Final Two-Way Comparison Study . . . 115
7.5.1 Method . . . 116
7.6 Chapter Discussion . . . 119
7.7 Conclusion . . . 122

8 Investigating Acceleration Curves in ASL 123

8.1 Related Work . . . 124
8.2 Modeling Acceleration Curves in ASL . . . 128
8.2.1 Research Question . . . 128
8.2.2 Method . . . 129
8.3 User Study . . . 133
8.3.1 Research Question . . . 133
8.3.2 Method . . . 134
8.4 Conclusion . . . 138

EPILOGUE FOR PART III 140

9 Conclusions and Future Work 141

9.1 Summary of Research Activities . . . 141
9.2 Contributions . . . 143
9.3 Limitations and Future Work . . . 146
9.4 Conclusion . . . 148

Bibliography 149

Appendices 166

A Appendix for Interview Study **167**

A.1 Simple of the Selected Stories for the 2008 Model and ASL-Speed Comparison . . 167

A.2 ASLSPEED2018 Survey Recruitment Flyer 171

A.3 ASLSPEED2018 Study Information Handout 172

A.4 ASLSPEED2018 Interview Demographic Paper for Participants 173

A.5 ASLSPEED2018 Interview Plan and Questions 176

B Appendix for ASL-Speed 2020 Study **182**

B.1 Simple of the Selected Stories 182

B.2 ASL-Speed 2020 Study Time Parameter Configuration 183

B.3 ASL-Speed 2020 Advertisement 187

B.4 Interview Plan and Questions 188

C Appendix for Other Contributions **196**

D Appendix for Annotation ELAN corpus **199**

E Publications **200**

List of Figures

1.1 Research focus of this dissertation. Starting from the left we build, extract, and clean human data till we get data ready for modeling. We engineer features and build Artificial intelligence (AI) models based on these features. Then, we build ASL animations and evaluate the quality of these animations. Furthermore, we conduct user studies to empirically investigate speed and timing in ASL with DHH users. Finally, we investigate the acceleration in ASL and present the summary of this work. 5

2.1 Movement-Hold model inspired from [95] 13

2.2 ASL generation pipeline with components labelled A-G as follows: (A) and (B) are data resources used by the pipeline: (A) dictionary storing individual ASL sign animations, (B) animation data for specific non-manual movements. This is followed by (C) creating a plan for an ASL sentence either by a human author or an automatic process (e.g. machine translation), with the output being (D) a symbolic script format. Next, the actual animation movements of a character must be specified based on this script, considering two types of factors: (E) non-linguistic issues that affect appearance and movement of human animations in general, and (F) ASL linguistic factors that influence aspect of the animation movement. The generated animation at the end of the pipeline is shown in (G) 16

2.3 This series of timeline images show how an ASL animation system may sequence decisions regarding speed and timing: (a) The input to the system is a script, specifying the identity of the words to use in the sentence - drawn from an internal lexicon in the ASL animation system and using a 'default' fundamental rate of signing. (b) Based on artificial intelligent modeling , the system must select where during the word sequence pauses should be inserted. (c) Next, the system differentially adjusts the rate of signing for each individual word, based on a variety of linguistic factors. (d) Finally, the system selects the duration of the pauses during the animation. This model of ASL speed and timing is used throughout this work. 19

3.1 Data splitting 37

4.1 User-interface of the ELAN annotation tool 48
4.2 Data extraction and pre-processing 52

5.1 Predictor screening report, with a red dotted line show which features were chosen for the modeling 71
5.2 Comparison of our new ASL-Speed model and the 2008 Model on the Pause Insertion task - for a subset of passages from [84] for which we added syntax annotation . . 77
5.3 Comparison among new ASL-Speed model and the 2008 Model - for Differential Rate and Pause Duration 78

6.1 Image of animation (left) seen participants in the user study, and transcript (right). Participants did not see the transcript 82

7.1 Visualization of five speed and timing parameters. The horizontal axis corresponds to a timeline representation of an ASL animation, with each rectangle representing the period of time of an individual word. (1) Variation in sign duration is illustrated in three alternatives A, B, and C, in which signs are produced at different speeds. (2) The transition time when the hands move from the final position of one sign and into the beginning position of the following sign may also be adjusted. (3) Signs may also be performed more quickly or more slowly due to various linguistic factors, which result in variation in differential signing rate, with more extreme speed-ups and slow-downs shown in the final timeline 3.C. (4) A signer may pause during signing for various linguistic reasons, and this timeline shows how the length of these pauses may vary. (5) The frequency with which someone may pause may also vary. 96

7.2 The interface for this study, which displayed five ASL animations of the same passage side-by-side, with each based on a different level of a particular timing parameter. Users indicated a scalar preference score (1 to 5) for each animation 100

7.3 A detail image of a zoomed-in region of Figure 7.2, showing one of the five ASL animation stimuli on the screen 101

7.4 User preference scores for five animations which varied in their average Sign Duration (in seconds). All pairwise differences are significant except between the pair marked with "n.s." 103

7.5 User preference scores for five animations which varied in their average Transition Time Between Signs (in seconds). Pairwise significant differences are marked with "**" ($p<0.01$) 104

7.6 User preference scores for five animations which varied in their average Pause Duration (in seconds). All pairwise differences are significant except between pairs marked with "n.s." 105

7.7 User preference scores for five animations which varied in their Pausing Frequency (represented as the percentage of inter-sign gaps where a pause occurs). All pairwise differences significant except one marked 106

7.8 The final velocity (vfinal) of a sign is based on its original velocity (voriginal) multiplied by a speed adjustment factor (Factor), which may be raised to some power (Exponent). In Figure 7, the values shown along the x-axis represent this Exponent 107

7.9 User preference scores for five animations which varied in the exponent used when applying their Differential Rate factor (see Equation 1). All pairwise differences significant except pair marked with "n.s.". The typical value for differential rate located at "1" which corresponds to the differential rate variability based on humans, with higher values reflecting more exaggerated variations in speed. 108

7.10 The interface for *Final Two-Way Comparison Study*, which displayed five ASL animations of the same passage side-by-side, with each based on a different level of a particular timing parameter. Participants subjectively evaluated each animation on a 5-point scale. 110

7.11 The five values selected for comparison in the Initial Five-Way Comparison Study (study 2 in the graph) and used as the basis for ASL animation stimuli for each parameter: sign duration, transition, differential rate, pause length, and pausing frequency. Participants evaluated each stimulus on a five-point scale, and the percentage of participants selecting each option (Very Bad, Bad, OK, Good, and Excellent) is shown, along with a divergent stacked bar graph redundantly visualizing these percentages. A mean score is also shown, calculated using 1 for Very Bad, 5 for Excellent, etc. Results of pairwise post hoc Wilcoxon tests comparing each level are presented, based on Bonferroni-corrected p-values ($p<0.05$), to explain how the "top two" values were identified for each parameter, which were subsequently compared in the Final Two-Way Comparison Study (study 3 in the graph). 113

7.12 A sample screenshot for one of the pairs of animations displayed in the *Final Two-Way Comparison Study*. 116

7.13 Participants' subjective ratings in *Final Two-Way Comparison Study*, which compared ASL animations with the top two levels from *Main Five-Way Comparison Study*, for each of the five timing parameters. The divergent stacked bar graph shows the percentage of participants who evaluated each animation as: Very Bad, Bad, OK, Good, or Excellent. For each parameter, significant pairwise differences are marked with * ($p<0.05$). 117

8.1 Velocity Curve for a Ballistic Movement 125

8.2 Velocity Curve for a Controlled Movement 125

8.3 These images showing velocity curves of hand movements during French Sign Language are reproduced from Duarte's dissertation [36]. Images (a-b) show movements within-signs, and (c-e) between signs. There were more examples of peak velocity occurring earlier for between-sign movements. The curves were standardized for time, so that they all had the same x duration (20 total x plot points); y values were left as-is for the purposes of maximum velocity comparison. The results are shown below. 127
8.4 Distribution of the time (normalized on a 0 to 1 scale) when the maximum velocity occurred during each movement section within ASL signs. 130
8.5 Distribution of the time (normalized on a 0 to 1 scale) when the maximum velocity occured during each movement section between ASL words (but excluding any between-sign sections at sentence boundaries). 130
8.6 Composite image of our previously shown results for within-sign movements in ASL, as compared to previously shown within-sign curves for LSF. 132
8.7 Composite image of our previously shown results for between-sign movements in ASL, as compared to previously shown within-sign curves for LSF. 132
8.8 A sample screenshot for one of the pairs of animations displayed in the user study. 135
8.9 Participants' average subjective preference scores for the baseline and the new animations shown in the study. 136

List of Tables

2.1 Timing in ASL and English 23

3.1 Related works on ASL animations 33

4.1 Types of used prompts. 42
4.3 Types of file in the Motion-Capture Corpus 44
4.5 Abstract non-manual annotations 51
4.7 Set of column used to build features 54
4.9 Complexity Index as four level of representation 55

5.1 List of predictor features used in this study, with a checkmark indicating if that feature used in each of our three models 66
5.3 Pause Prediction Model Results. The Decision Tree and SVM classifiers were implemented in MATLAB using the Classifier Learner Package, while the Linear-Chain CRF classifier was implemented using the sklearn-crfsuite package in Python. Parameter setting for the models are share in the footnote 72
5.4 Differential Signing Rate prediction model results 74
5.5 Pause Duration prediction model results 75

B.1 Sign duration values 184
B.3 Transition values 184
B.5 Pause duration values 185
B.7 Pausing frequency values 185
B.9 Differential rate values 186

Chapter 1

Introduction

1.1 Motivation for Research on Synthesizing ASL Animation

The World Federation of the Deaf has reported that there are more than three hundred sign language around the world [112]. These sign languages are used by 70 million deaf people. In particular, American Sign Language (ASL) is used as a primary means of communication for more than half million people the United States [104, 112]. ASL is a complete natural language, which uses movements of the hands, body, head, and facial expression to convey linguistic structure. ASL a different language than English, and contrary to popular misconceptions, it is not just a simple representation of an English language sentence using the body. ASL, similar to most other natural language, has its own syntax structure, word order, and lexicon which is different from spoken and written English.

While on one hand there are many Deaf or Hard-of-Hearing DHH[1] individuals with excellent English literacy, on the other hand, some DHH individuals who experienced a low level of language-exposure during their childhood (and some educational circumstances) may have a lower level of English language literacy. As a matter of fact, standardized testing has shown that the majority of deaf secondary-school graduates (median age is 18 years old) in the United States have English

[1] there is variation in the meaning of the terms:Deaf, deaf, and DHH for consistency in this document I will use DHH only

reading level at the fourth-grade level (typical of U.S. students of age of ten) [137]. Because of the linguistic differences between the two languages (ASL and English), there are many individuals who are fluent in American Sign Language but have many difficulties reading English text.

As a result of these literacy differences, it is understandable that a barrier faced by many deaf adults in the U.S. is accessing information in the form of English text, including on websites [115]. While it is common for companies, organizations, or governments to offer versions of their website in multiple written languages, it is much less common to see websites that provide information content in the form of sign language. ASL does not have a written script that is commonly used among the community who uses the language; so, it is not possible to provide some written form of the language on websites. A seemingly simple solution for providing ASL online would be to post videos of a human signer (who performs ASL) onto websites, but this is actually not a practical solution due to challenges faced in producing, managing, and updating video recordings of humans performing sign language [65]. As compared to the ease of updating a text file containing a written language, updating information in a video of sign language requires hiring a fluent signer and re-recording information content. An alternative solution is to provide computer animation (carefully produced by a professional animation expert) of sign language on websites, but the time and resources needed for human animators to carefully produce fluent and understandable motions of a 3D virtual human character are also substantial [65]. Prior research [2, 38, 60, 63, 126] on building American Sign Language animations has investigated how to (partially) automate the creation of such animations, to reduce the time and skills needed by the human who is authoring the message. Rather than a computer animator, a human who is skilled in ASL could author a symbolic representation of the message, which software could synthesize into a computer animation.

While the author of such an input script must be someone who is fluent in ASL (ideally a Deaf individual with native fluency in ASL who could translate the English text), we do not believe it is reasonable to burden the human author with making subtle numerical decisions about, e.g., the number of milliseconds to pause between words or the specific speed variations of individual words throughout a longer discourse. While fluent signers may be able to produce ASL (using their body) that contains these natural timing subtleties themselves, they may not be able to choose specific numerical values for all of these properties, for all words in a script for generating a computer animation.

1.2 Focus of This Dissertation

In prior work, our lab has collected a dataset called the "ASL Motion-Capture Corpus" [100], which contains video and motion-capture recordings of fluent ASL signers, with linguistic annotation by experts. This annotation includes timing information for when particular words or linguistic structures occur during the signing. Prior to this dissertation research, our lab had used only some small portions of this resource to investigate, using data driven approaches, different aspects of ASL animation: inflecting verb movement [97], facial expression [83], and spatial reference point locations [73]. Given the success of these prior projects at using motion-capture recordings to build models of how human ASL signers behave, this dissertation research includes three main goals:

1. We built an updated dataset by adapting and enhancing our existing corpus and adding new layers of linguistic annotation to support research on data-driven modeling of ASL speed and timing. We documented the workflow for speed and timing modeling from this resource to generalize the use of this corpus for other speed and timing research.

2. Next, given a sequence of ASL words in a message (along with a limited amount of additional structural information provided by the human author, e.g. sentence boundaries), we used our new dataset to create models that can automatically identify the speed and timing values for each individual word in the message. Finally, we evaluate these models using dataset-based and user-based studies with DHH participants.

3. We conducted several empirical studies to investigate speed and timing preferences among DHH ASL signers with native fluency, with two aims: (a) to identify ASL signers' preferred values for speed and timing parameters and (b) to determine whether participants prefer animations with timing values that differ from those in typical human signing. Finally, we analysed the acceleration curves in the recordings of ASL human signers from our motion-capture corpus, and we conducted a user-based evaluation with fluent ASL signers to investigate viewers' preferences between animations with simplistic uniform acceleration during movements, as compared to animations with acceleration curves based on human recordings.

While these subtle timing values may seem negligible at first blush, prior perceptual studies at our lab [60] of the quality of sign language animation (conducted with DHH participants) has revealed that minor variations in minor timing parameters can lead to **significant differences** in

users' perception of the quality, and this research has found that even tiny errors in these parameters will lead to less understandable animations [63]. Thus, we investigated how to automatically identify the speed and timing concerns for each word in an input ASL script, by training machine-learning models on movement data from human ASL signers. Our ultimate goal is to automate this aspect of animation synthesis and to create understandable and realistic ASL animation with minimum human effort.

Our evaluation methodologies consist of both dataset-based and user-based studies: *(1)* Our dataset-based evaluation uses a cross-validation approach to compare a prior state-of-the-art model of speed and timing for ASL animations, which had been rule-based [60, 63]. This baseline model is compared to our new data-driven model. *(2)* For user-based evaluation we generate ASL animations using our modeling approaches and conduct a study with DHH participants to compare animations side-by-side (one based on this new model and the other a baseline timing model). Participants offer their subjective assessment of the animation quality and recommendations for future improvements.

While this dissertation focuses on the U.S. and ASL, there are similar motivations for research on automatically synthesizing sign language (and a need for high-quality speed and timing in such animations) among users of French Sign Language (langue des signes française, LSF) [126], British Sign language (BSL) [31], Turkish Sign Language (Türk İşaret Dili, TİD) [131], Arabic Sign Language (ArSL) [13], etc.

Figure 1.1 illustrates the overall methodology of our data-driven modeling research in this dissertation. The key premise of this work is that to improve ASL animation quality, we should consider quantitatively how humans move when they produce ASL. Specifically, we should use motion-capture data and data-science analysis techniques, to produce guidance for predictions as to how to generate ASL animation. However, training and evaluating on datasets is not enough - We must also consider whether Deaf ASL signers actually prefer animations based on these predictions, and we should consider whether the movement and timing values derived from humans may need to be adjusted when applied to computer animations of ASL (which perhaps should occur at slower speed or with other differences preferred by viewers of those animations).

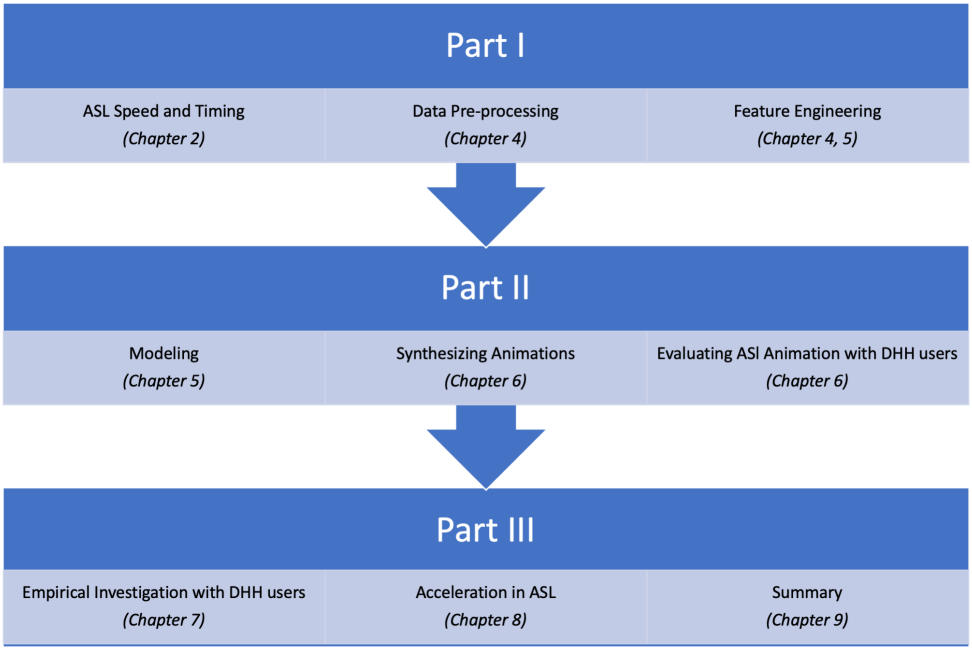


Figure 1.1: Research focus of this dissertation. Starting from the left we build, extract, and clean human data till we get data ready for modeling. We engineer features and build Artificial intelligence (AI) models based on these features. Then, we build ASL animations and evaluate the quality of these animations. Furthermore, we conduct user studies to empirically investigate speed and timing in ASL with DHH users. Finally, we investigate the acceleration in ASL and present the summary of this work.

1.3 Contributions of This Dissertation

The contributions of this dissertation research include:

Contribution 1: We have created a new American Sign Language Speed and Timing Dataset, which is an enhancement to our lab's pre-existing motion-capture corpus of ASL. As part of this work, we transferred our prior motion-capture corpus to a new linguistic annotation platform that has become standard among sign-language linguistic researchers, ELAN [130]. We have added layers of annotations and document our data preprocessing procedures which were necessary to make this resource useful

for speed and timing research. We have also documented our feature engineering process to create input for machine-learning modeling, so that it is easier for future researchers to work with this new dataset.

Chapter 4 will discuss Contribution 1.

Contribution 2: We empirically determined which features[2] were most influential in the speed and timing prediction models, e.g. via a feature-ablation analysis. Since our goal is to build a system that could convert a script that specifies an ASL message into an animation automatically, it is useful to identify a minimal set of information that the person writing the script must specify in order for our software to operate. We performed this analysis for each of the following three modeling tasks:

- **2.A:** Empirically determine the best subset of features needed to be used for building a predictive model for predication the **prosodic break (a pause)** after each word.
- **2.B:** Empirically determine the best subset of features needed to be used for prediction the **time-duration of this break/pause**.
- **2.C:** We empirically determine the best subset of features needed to be used for modeling the **variation of the speed for each particular word** in the message.

Chapter 5 will discuss Contribution 2.

Contribution 3: We empirically determined whether a machine-learning modeling trained on a final subset of the linguistic features out-performs prior state-of-the-art rule-based approaches for the task of predicting the timing parameters for ASL multi-sentence passages. Specifically, in a cross-validation analysis of held-out data, we automatically identified the following three speed and timing values for each individual word in a message:

[2] A feature, in machine learning, is an individual measurable property or characteristic of a phenomenon [20]

3.A: Is there a**prosodic break (a pause)** after this specific word? ASL signers will naturally pause at various locations during a message, typically more frequently at structural boundaries, e.g. as discussed in [118].

3.B: If so, what is the **time-duration of this break/pause**? ASL signers are also more likely to use longer pauses at more important structural boundaries [53].

3.C: Given the overall signing rate that we seek to produce, what is the **variation of this speed (slightly faster, slightly slower) for each particular word** in the message? ASL signers will generally slow down at the end of sentences, or change their signing speed for individual words, for a variety of reasons [52, 143].

Chapter 5 will discuss Contribution 3.

Contribution 4: Empirically determine whether Deaf ASL signers prefer animations of multi-sentence ASL passages in which timing values are determined by these new models or by the previous state-of-the-art rule-based technique.

Chapter 6 will discuss Contribution 4.

Contribution 5: There is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation - for instance, prior work had found that users may prefer animations to be slower than human videos [63, 69]. Thus, for each of the timing parameters for ASL animation, we empirically determined which values of that parameter are preferred by Deaf ASL signers via an experimental study, in which animations with a range of such values are displayed for comparison.

Chapter 7 will discuss Contribution 5.

Contribution 6: Prior research on speed and timing of sign-language animation has not specifically investigated the issue of predicting acceleration curves for the movements of the character's body [7, 63, 69]. Further, some prior linguistic research has observed different classes of acceleration curves used during or between words in French Sign Language [36], but such an investigation has not been performed for ASL.

Thus, we conducted an analysis on our new dataset of motion-capture patterns of human movements, to empirically determine whether there are common categories of acceleration curves present in different linguistic environments, e.g. within ASL signs, between ASL signs, or near sentence boundaries.

Chapter 8 will discuss Contribution 6.

Contribution 7: Following the same logic as for Contribution 5 above, since there is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation, we empirically determined whether accuracy in the use of particular acceleration curves influences the subjective judgements of Deaf ASL signers, as to the quality of the ASL animation.

Chapter 8 will discuss Contribution 7.

1.4 Structure of This Dissertation

To provide the readers with essential background knowledge, Chapter 2 will provide a linguistic background on ASL timing technology, explaining the speed and timing concepts which are used within the later chapters of dissertation. Chapter 3 will discuss in depth prior research on speed and timing for sign-language animation, including evaluation approaches. This document will then begin to discuss my research work, with Chapter 4 focusing on our use of motion-capture data and dataset pre-processing steps. Chapter 5 presents our machine learning modeling approach, feature selection, computational linguistic cross evaluation, and proposed baseline to select the robust model. Chapter 6 will focus on evaluation mechanisms including the description of the human evaluation process, user study stimuli and procedure, and the feedback from users. Chapter 7 focuses on user based investigation of speed and timing parameters in ASL and compares user' preferences between ASL animation and humans. Chapter 8 presents our investigation on acceleration curves in ASL. Finally, a summary of the contribution will be presented in Chapter 9.

Part I: Understanding The Generation Of ASL Animations And Dataset Creation

PROLOGUE TO PART I

In Part I, Chapter 2 will begin by introducing the reader to important definitions related to speed and timing in human biomechanics. We will discuss the speed and pauses in ASL; furthermore, we will compare these concepts of speed and pausing in spoken English language and in ASL. Next, we will discuss the importance of the speed and timing parameters in the process of synthesizing ASL animation.

Chapter 3 will discuss prior work related to generated animations of American Sign Language. Specifically, we will discuss prior computational linguistic work in synthesis, using both rule-based and data-driven approaches. Then we will discuss some analogous research in the field of speech synthesis. Finally, we will discuss the literature related to speed and timing in sign-language animations, and we will review how prior research has evaluated the quality of these animations to determine if they are acceptable and useful for the DHH community.

Finally, Chapter 4 will explain our process for enhancing a motion-capture corpus of ASL, which our lab had collected in prior work, to create a dataset that can support data-driven research on ASL speed and timing. The discussion of data extraction and cleaning in this chapter will focus on transferring the original corpus to the ELAN annotation platform, adding new layers of linguistic annotation, extracting linguistic features from the corpus, and the process of annotating additional ASL recordings. The aim of that work is to create a new dataset that can be used for training predictive models for sign language animations. Two key contributions from this work include our documentation of how we processed a corpus to make it suitable for speed and timing research, and the actual enhanced version of the ASL motion-capture corpus from our laboratory, which will be disseminated in this dissertation research.

Chapter 2

Key Concepts for ASL Speed and Timing

This chapter will provide essential background information and terminology that will be used throughout this dissertation document. This chapter is organized as follows: Section 2.1 provides some key background for sign language linguistic and animation representations. Section 2.2 will present the standard pipeline model for creating sign language animations. Section 2.3 will introduce key terminology about speed, acceleration, and the biomechanics of human body movement, with some discussion about how these concepts related to the generation of ASL. Section 2.4 will begin with an explanation how speed can be parameterized into several components in ASL, and several important terms will be defined for these various parameters, which will be used throughout this document. Later, Section 2.5 will focus on key concepts that underlie the insertion of pauses during ASL signing and the duration of those pauses. Both Section 2.4 and Section 2.5 describe concepts which are the focus of prediction models in this dissertation research, and those sections will also provide readers with sufficient linguistic background to understand the most important aspects of this dissertation work. Section 2.5 will provide evidence that accurately selecting these timing parameters is important when producing ASL animations, to provide motivation for this research. In addition to providing key concepts that relate to the research, this chapter will also

enable the reader to more easily understand elements of the literature survey discussion, which will appear in Chapter 3.

2.1 Sequential Representations of ASL Signs and Animation

As discussed previously, American Sign Language (ASL) is a natural language that is articulated using movements of the eyes, face, head, torso, arms, and hands. Although the language may include some complex use of 3D space [48], facial expressions used to convey syntactic information or other meaning [82], and other complex phenomena, most ASL signing consists of sequences of individual signs (words), performed using the hands/arms, which are assembled into longer phrases or sentences. Traditionally, individual ASL signs are thought of as consisting of combinations of handshapes, 3D orientations of the hand, locations of the hand, and movements through space [23, 90]. Various sign-language phonological researchers have proposed ways of representing the structure, over time, of individual words. For instance, there have been multiple linguistic representations of the time structure of ASL signs, including the Movement Hold model of Liddell and Johnson (1989) [95], shown in Figure 2.2, the Hand Tier model (Sandler, 1989) [121] , the Moraic model (Perlmutter 1992) [117], and other recent in this area. A common feature of these models of sign language phonology is that they represent a sign as a combination of static positions (moments in time, where the hands have a particular handshape, orientation, and location) along with some "movements/transitions" in-between these moments. Many of these models also discuss how when multiple signs are assembled together into sentences, then the transitional time between the end of one sign and the beginning of the next may be considered a type of "movement/transition" between the two words. Notably, some of these static positions may be instantaneous movements of the hands through a particular target location, and at other times, the hands may actually hold/pause at these locations for some duration of time.

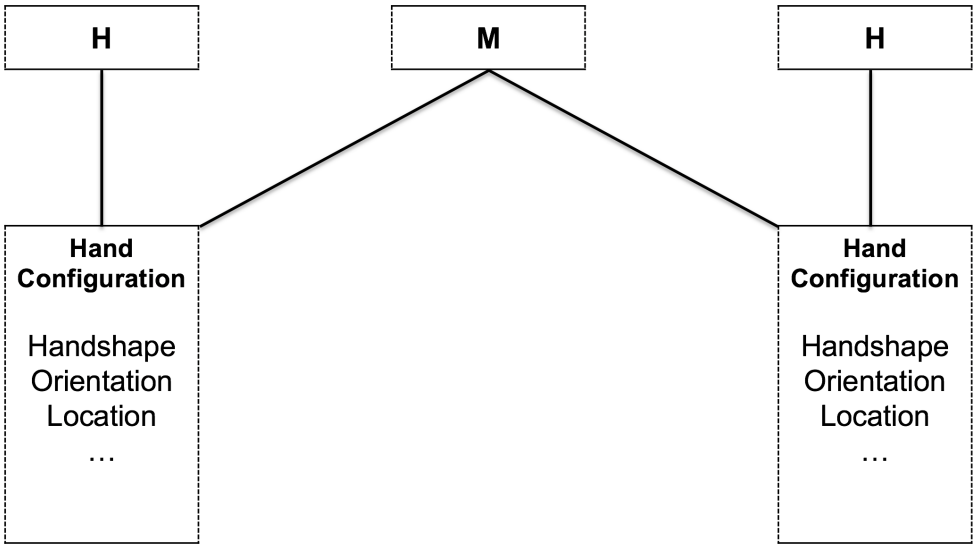


Figure 2.1: Movement-Hold model inspired from [95]

In a similar way, many researchers who have been interested in producing animations of sign language have considered representations of the language that also consist of key moments in time (when the hand is at some instantaneous "snapshot" of handshape, orientation, and location in space), with some movements/transitions between these locations. Such an approach follows a classic method of computer animation creation that is commonly referred to as "key frame animation," which consists of setting key moments on a timeline where a particular configuration of an animated character is posed (a concept inherited from professional Walt Disney animators who drew the key frames), with transitional frames drawn by assistants [16,107,108,123]. A key-frame based method of specifying sign language animation has been used in previous work from our laboratory [48,63], as well as successful commercial software that can be used for producing sign language animation [133].

An alternative to the key-frame approach would be to specify the movements of a sign-language character by using motion-capture technology to digitize full recordings of the 3D movement of a human, for each word. Motion capture based animation the movements from real human, or analysis of videos of humans videos [4], with the resulting data mapped to a to animation avatar [35].

A third approach for how to animate a character in computer animation or games would be procedural animation, in which characters' movement is based on algorithms or procedures, e.g. [15,42,59,123,129,144]. As discussed in [27], an advantage of keyframe-based animation is that it

supports greater flexibility in the animation system, which can more easily blend from one sign into the next, or modify elements of the performance. Given the focus on keyframe-based animation in recent work on sign-language animation, as well as the existing context of our laboratory's prior research in this area, for this dissertation research, a keyframe animation method is used. Thus, we represent ASL signs (sequences of movements and holds) as sequences of keyframes (with transitional movements between them). This concept can be utilized not only for representing single signs, but it can also be a framework for representing sequences of signs during entire sentences or phrases.

2.2 General Pipeline for Sign Language Generation

A standard pipeline model for creating sign language animations is presented in a recent survey of the field of sign language generation and translation [25], which was the best paper award winner in ASSETS '19. As shown in Figure 2.2, this pipeline consists of multiple steps, each of which focus on a different stage of sign-language synthesis. In sign-language animation systems, there is generally a lexicon (a dictionary with a collection of individual words in ASL), which can be used to create new sentences or longer messages. Also, there are some non-manual components to represent other aspects of animations, like grammatical or emotional information. When someone wants to produce a sign language animation, they can arrange the animated signs and the non-manual components using one of two methods: either (a) a person manually edits the plan (assembling words on a timeline) for a sentence using some authoring software [18,24] or (b) software automatically plans the message, e.g. using machine translation technology [41,66], to create a script for the message that consists of these individual dictionary items. This plan is generally a symbolic representation [2,24,40] of an easy-to-update script that describes how to assemble signs and non-manual elements into sentences. Notably, there is no common standard for the script formalism used by different researchers to plan sign-language animations.

When an animation is to be produced, then the "animation content" of each of these signs is loaded from the system's dictionary. This animation content consists of some sequence of key frames, specifying targets (at particular moments in time) for the hand location, orientation, and shape. These animation specifications for individual words are then strung together in a sequence to produce an overall sentence or message. The animation content specifies the appearance of the generated avatar. For example, this may consist of fine-grained movement details based on the

biomechanics of the body motion to move between keyframes, to avoid collisions, and to optimize naturalness of the animations. Often this animation stage of movement planning will be performed by a human animation engine, such as [31, 133]. This final animation stage of the pipeline will also specify how the animation will look, for example lighting, shadows, and textures. And the animation pipeline may also include additional biological behaviors, e.g. breathing and blinking.

Although there has been significant research in the field of computer animation about producing natural animations of humans, e.g. with much work in the field of video gaming [46, 105, 138, 145], there are also some ***ASL-specific issues***. These ASL-specific issues include important aspects of generating animation that are particular to *sign-language* animation generation, and these issues include some subtle yet important aspects of sign language generation. For example, prior research [127] has discussed animations that include *coarticulation effects*, i.e. when the location or shape of the signer's hand for the end of one sign is altered relative to the details of the next sign in the sentence. *Facial expressions* are another subtle aspect of sign language motion planning which has already been deeply investigated by Kacorri in [82], who conducted multiple user studies with deaf participants to study this aspect. Other ASL-specific issues in animation motion planning include *speed*, *timing*, and *acceleration*, which are influenced by various ASL linguistic properties, yet these three issues have not been the focus of much prior research.

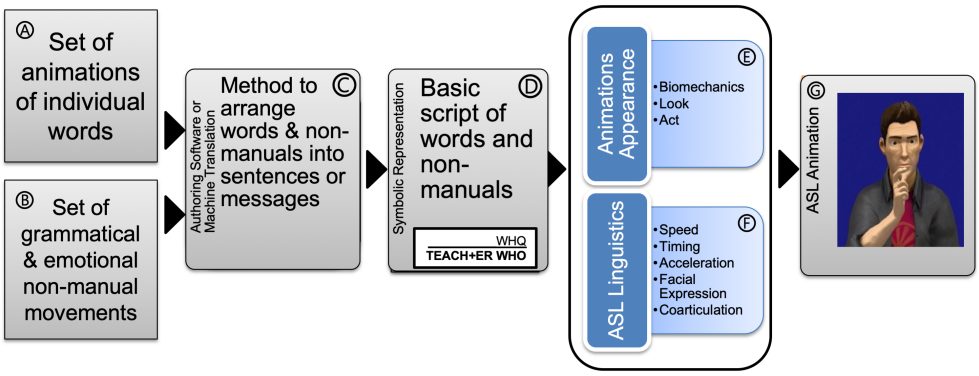


Figure 2.2: ASL generation pipeline with components labelled A-G as follows: (A) and (B) are data resources used by the pipeline: (A) dictionary storing individual ASL sign animations, (B) animation data for specific non-manual movements. This is followed by (C) creating a plan for an ASL sentence either by a human author or an automatic process (e.g. machine translation), with the output being (D) a symbolic script format. Next, the actual animation movements of a character must be specified based on this script, considering two types of factors: (E) non-linguistic issues that affect appearance and movement of human animations in general, and (F) ASL linguistic factors that influence aspect of the animation movement. The generated animation at the end of the pipeline is shown in (G)

As shown in Figure 2.2, the research focus of this dissertation is in the area labelled as "ASL Linguistics" (section F of the Figure 2.2). As discussed in Section 2.3, after a "basic script" of an ASL sentence has been assembled (which makes reference to specific items in the system's lexicon of signs), it is still necessary to select various speed and timing parameters to convert this script into an animation. In addition to blending in-between the keyframes of the animation information for each sign in the script (i.e. "connecting the dots" between the targets for how the hands should move through space), it is also necessary to specify particular speed and timing for each portion of the animation.

For example, in my research, I have investigated some ASL-specific animation planning issues which include pausing [12] as well as speed and timing [7]. I have also investigated timing parameters via user studies with DHH participants [9, 11]. In these user studies, I have modified some linguistic timing parameters of human-authored "basic scripts" of ASL animations (using the

Vcom3D software platform) to create new ASL animations with various timing parameters, which participants evaluated.[3]

2.3 Speed, Acceleration, and Biomechanics of Hand Movement

For clarity, we provide definitions of some key physics concepts used throughout this work. Although sign language consists of movements of the head and torso, this dissertation research primarily focuses on the animation of the arms and hands of a virtual character performing ASL. For this reason, the discussion of these concepts below focuses how the hands of a human may move through space, during ASL.

Speed refers to the rate of change of distance with time. Thus, speed is a directional scalar quantity that specifies how fast (or slow) a hand is moving. It is the rate of change of displacement with respect to time [57, 147] (velocity is speed in a given direction). In other words, velocity is speed in a given direction. Speed are measured in units of distance divided by time, e.g. centimeters per second.

Acceleration refers to the rate of change of velocity (speed) with respect to time. Acceleration could be positive, negative, or could have zero value [57, 147]. From an ASL perspective, during positive acceleration of the hand, a signer would be increasing the speed of their hand, e.g. perhaps when they are beginning to move the hand after it has been stationary for some period of time. A negative acceleration would indicate that the signer is decreasing the speed of their hand, e.g. if they are slowing down their motion of the hand through space. If the human is moving their hand through space at a constant speed, then there is zero acceleration during this time.

From Newton's laws of inertia and acceleration, we know that effort/energy must be exerted to change the velocity of something, and the amount of effort/energy affects the amount of acceleration [2]. Elements of the human body moving through space are also subject to these laws of physics, i.e. a human must exert some muscle effort to increase the acceleration of the hands through space.

[3] While not the main focus of this dissertation, this author has also conducted and supervised other research projects investigating the modeling of the placement of spatial reference points (locations where items under discussion are placed in the space around a signer for subsequent pronoun reference) [48] and investigating the user-interface design for authoring basic-scripts of ASL sentences [89] or the movements of individuals signs [76].

The influence of gravity and the sudden slow-down in motion when the hands collide with other parts of the body also have significant effects on acceleration values for the hands. Research in the field of biomechanics has studied the complex relationships between human muscle energy, bone structure, and comfort for various poses and movements of the human body [91].

Significant research in the field of computer animation that has sought to incorporate sophisticated models of human biomechanics to produce realistic movements and timing of the human body through space. In fact, some researchers have looked how to control speed and timing in animations to adjust the appropriate values for speed and timing between animation keyframes e.g. [80,94,96,141]. The authors in [141] found that timing and speed are very important elements of animation because the movement and the speed of character reflects the weight and the size of the character [141]. According to Frank and Ollie's book *"The illusion of life: Disney animation,"* emotions (relaxed, nervous, and excited) of the character are conveyed more by the character's movements than by the character's appearance [80, 94]. Furthermore, some researchers have focused specifically on how arms move with realistic speed and timing, e.g. for tasks like grabbing or reaching for objects [28, 128]. This prior research that is not ASL-specific (i.e. from a biomechanics or physics perspective) can still influence the movement of an animated ASL virtual human, as shown in section E in Figure 2.2. In fact, many researchers in the field of sign language animation have built their systems for producing animation of virtual humans using underlying animation engines for human animation that consider these non-ASL-specific factors. For instance, the animation pipeline used in prior research at our laboratory has been based on a human-animation platform that includes consideration of these biomechanical factors as the final phase of the animation planning process [133]. As more important/worthy than biomechanics and physics, the sign-language linguistics literature has established that speed and timing of movement during ASL is also based on key linguistic factors, as discussed in the next section.

2.4 Speed and Timing in ASL

There are multiple factors that may motivate variations in the speed and timing of a human while they are performing ASL signing, including various prosodic factors. The term *prosody* generally refers to how humans convey grouping or prominence of linguistic units through variations in a language performance [33], and there has been significant linguistic research on how elements of spoken and sign language can indicate such information, e.g. [34, 44, 120, 122]. For spoken

languages, researchers have examined how changes in timing, volume, and other aspects of human speech may convey important prosodic information about a message [44, 140]. There has also been significant research on prosody for sign languages, e.g. [34, 120, 122], which has examined how changes in body movement, eye-aperture, lengthening of signs, pausing, or other factors can convey prosodic information. As a simple example, ASL signers are known to often include a subtle pause in-between important syntactic phrases or sentences [53, 55], and it is known that ASL signers typically reduce their speed as they approach the end of such phrases [53, 55]. This section discuss how the speed and timing parameters for sign-language may be conceptualized, and it defines five key parameters (shown in Figure 2.3) which will be discussed throughout this dissertation research, namely: fundamental rate, base duration, differential signing rate, pause insertion, and pause duration.

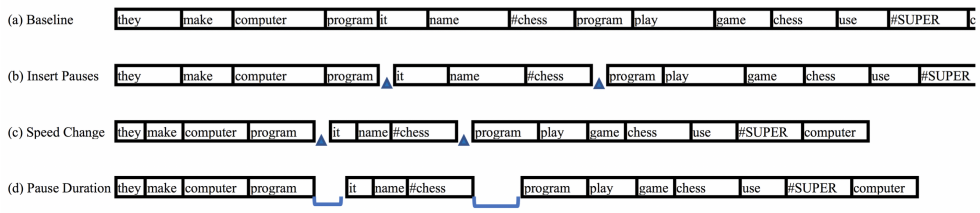


Figure 2.3: This series of timeline images show how an ASL animation system may sequence decisions regarding speed and timing: (a) The input to the system is a script, specifying the identity of the words to use in the sentence - drawn from an internal lexicon in the ASL animation system and using a 'default' fundamental rate of signing. (b) Based on artificial intelligent modeling , the system must select where during the word sequence pauses should be inserted. (c) Next, the system differentially adjusts the rate of signing for each individual word, based on a variety of linguistic factors. (d) Finally, the system selects the duration of the pauses during the animation. This model of ASL speed and timing is used throughout this work.

In this dissertation research, we focus specifically on how aspects of speed and timing of ASL can be predicted, to support the generation of more realistic and understandable animations of ASL, which may reflect some of these typical patterns in human ASL signing.

2.4.1 Fundamental Rate (Words Per Second, Overall)

Some signers, in some settings, are faster in their personal rate of signing than others. In a similar way, some speakers are "faster talkers." In this work, we refer to this overall rate of signing communication as the signer's fundamental speed, and it can be expressed as the number of "words per second" that the signer tends to produce, as averaged over some longer span of language production. As we mentioned above, the input to the system is a script, specifying the identity of the words to use in the sentence - drawn from an internal lexicon in the ASL animation system and using a 'default' fundamental rate of signing. Prior **linguistics** researchers have analyzed videos of human signers to determine that typical fundamental rates of signing vary between 1.5 and 2.37 words per second, depending on context [19, 43, 51, 52]. Prior research studies focused on ASL animation [63, 69], have found that humans viewing animations tend to prefer fundamental signing rates ranging from 0.9 to 1.2 signs per second. In general, Fischer et al. [43] found that there was remarkable drop of in human's comprehension of videos if the sign language video played faster than 2.5 times of its normal speed. In comparative studies, researchers found that ASL signing conveys information content at an equivalent rate as spoken English [19]. While ASL fundamental rate is slower (fewer words per second), ASL sentences tend to be shorter (consisting of fewer individual words per sentence) to convey the same ideas. Although ASL fun fundamental rate is slower, ASL convey information at the same rate as spoken English.

2.4.2 Base Duration (of a Word in a Lexicon)

Each ASL sign consists of a particular number of movements of the hands, to particular locations in space or on the body. As such, ASL signs have a basic duration that varies from word to word, with some signs naturally taking longer to perform than others, since some ASL signs consist of a more complex or larger set of movements. Loosely, the reader may consider an analogy to spoken languages, in which spoken words vary in the number of syllables they contain (e.g., "so" vs. "consequentially"). Some **linguistics** researchers have conducted work on lexicography of sign languages, in which they assemble lexicons (vocabulary lists or dictionaries) for various sign languages, often based on identifying large numbers of examples of a particular word used in multiple video recordings of sign language that have been analyzed, e.g. [139]. If enough examples can be identified for a particular word, researchers can estimate a typical range for the base duration of that word.

In this dissertation research, we define the "base duration" of an animated ASL sign as the time-duration of the entry for this sign in the animation system's dictionary, as it had been engineered by the animator who authored that sign animation. When considering a specific instance of an ASL sign in our motion-capture dataset, we consider the base duration as the span of time that the linguistic annotator had marked when annotating the corpus.

2.4.3 Differential Signing Rate (of a Word as Performed)

During a sentence or some longer span of signing, a signer will vary in the differential speeds at which each individual word is performed. For instance, as a signer approaches the end of a sentence, the final word(s) may be performed more slowly. Thus, the differential signing rate is a property of an individual word, as performed in an individual instance - such that the sign may be performed more quickly or more slowly, for various contextual reasons. Notably, the specific amount of time that a signer uses to perform some word, on a particular occasion, is based on an interaction of all three of these factors above (fundamental rate, base duration, and differential rate). Prior **linguistics** researchers have investigated the speed of individual words in an ASL performance, by analyzing video recordings of human ASL signers [52, 143]. There is a challenge in studying differential rate from recordings: When observing the amount of time someone requires to perform an ASL sign in a particular instance, the duration is based on the overall fundamental rate and the base duration of the word. Unless the researchers are able to calculate average speed for individual signers and average duration of a particular words, it is difficult to isolate the differential rate of a particular word in that performance.

In this dissertation research, we view differential signing rate as a factor that can be applied to the timing of a sign. Specifically, when this factor is greater than 1, then there would be two effects on the sign: The speed of the hands during "movements" would be increased, and the amount of time the hands spend stationary during "holds" is reduced.

2.5 Pauses in ASL

In addition to modulations in the speed of signing (according to the three parameters defined above), human ASL signers will sometimes pause in-between words. This section will discuss

various parameters relevant to such inter-sign pausing in ASL, namely: determining where to insert pauses in an ASL message and selecting the duration of these inserted pauses.

2.5.1 Pause Insertion

Prior linguistics researchers have studied where ASL signers tend to pause, e.g. Grosjean et al. [53] based their analysis on observations of video recordings of native ASL signers. They conceptualized the decision as to whether a pause should occur at a particular word boundary as being binary in nature, i.e. they determined that at any word-to-word boundary, there either exists or does not exist a pause at that location. (The reader may note that an alternative conceptualization would be to assume there is some amount of "pause" at every boundary - sometimes with duration 0, meaning that there is no apparent pause). In our work, we adopt the binary view of Grosjean at el. [53], during their analysis of videos, Grosjean et al. [53] determined that there were pauses at approximately 25% of word boundaries in an ASL passage. They also proposed a scheme, based on a sentence's syntactic structure, for insertion of pauses based on the syntax structure of the sentence and proximity of a word boundary to other nearby pauses.

2.5.2 Pause Duration

In their analysis of ASL videos of signing, Grosjean et al. [53] noted the duration of each pause they observed, and they calculated average pause duration, when the pause was at various key locations: 229 milliseconds between sentences, 134 milliseconds between conjoined sentences, 106 milliseconds between noun or verb phrases, 11 milliseconds if within verb phrases, etc. The main take-away from the work of Grosjean and Lane was that pause duration in ASL was related to the syntactic structure of the sentence; they also proposed some formulas for timing parameters, based on a sentence's syntactic structure. Grosjean and other researchers proposed a scheme, based on a sentence's syntactic structure, for predicting the duration of pauses at various syntactic locations [53], which was utilized in a prior rule-based ASL speed and timing model [63], which is the current state-of-the-art method for predicting speed and timing in ASL animation. The linguistic findings about pause duration have inspired the selection of features for our models, as discussed in Chapter 5.

Table 2.1: Timing in ASL and English

Factor	Spoken English	ASL
Pause length at sentence boundaries	445 ms	229 ms
Verb phrases and conjoined sentences	From 245 ms to 445 ms	134 ms
Within phrasal constituents	<245 ms	<106 ms

2.6 Typical Speed and Timing Values for ASL and for Spoken English

Prior psycholinguistic studies have focused, via different studies, on the timing and pausing of ASL and spoken English [51, 53, 55]. For example, some of this research has included comparative studies, between English and ASL, yielding numerical timing parameters for each language. In spoken English, Grosjean and Lane [55] found that longer pauses take place at sentence boundaries (pause length longer than 445 ms); shorter pauses take place between noun phrases, verb phrases and conjoined sentences (pause length range between 245 and 445 ms), and the shortest pauses occurred within phrasal constituents (pause length less than 245 ms) [2, 143]. Table 2.1 summarize a numeric comparison for pauses length in spoken English and ASL.

2.7 Why Speed and Timing Parameters are Important for Animators

In prior research, our lab had conducted an experimental study to understand the effect of modifying the speed and timing parameters of ASL animations. In this study, our lab had created a rule-based system[4] for predicting these speed and timing values [52, 53, 60, 63], and then the lab generated ASL animations with various speed and timing values, to compare alternative versions of the system. In an experimental study, participants who were native ASL signers evaluated the quality of ASL animations. Participants indicated their subjective impression of the understandability and naturalness of animations on a numerical scale (e.g. 1-to-10 Likert scale used in [63]), and they answered some

[4]Additional details about the rule-based algorithm used in this prior research is discussed in greater detail in Section 3.3, but at the moment, this prior research is only mentioned here to discuss how humans who view ASL animations are sensitive to changes in these parameters.

comprehension questions about the stories shown in the animations [63]. This study verified that (a) inserting pauses at linguistically motivated locations in animations of ASL and (b) modifying the duration of ASL signs based on their location within a sentence, both led to statistically significant improvements in the understandability of the animations. When animations did not insert pauses or adjust sign durations in this manner, the animations were significantly less understandable to native-ASL-signer participants [60, 63].

In addition, this prior work [63] established that participants prefer for ASL animation to be displayed at a particular overall speed that is slower than the speed of human ASL signers in videos. Specifically, the study [63] found that participants preferred animations in the range of 0.9 - 1.2 signs per second, with the authors speculating that an ideal speed value for ASL animations may be around 1.1 signs per second. Again, when the speed values were farther from this ideal value, researchers in [63] observed both lower subjective scores from participants as well as lower comprehension question response accuracy.

In some recent preliminary research [7], we conducted an interview-based study with DHH participants who were native ASL signers; we showed the participants some examples of ASL animations with variations in their speed and timing values. The open-ended comments from participants further suggest the importance of having accurate speed and timing in ASL animations. For instance, in response to the question "Do you think pauses and speed are important?" participants replied:

- "I prefer signing that has normal pace and normal pause. I am worried about them being too fast or too slow. The speed needs to be stable for everyone to understand. The pausing is important. For example, when signing for a long time, I lose information because the animation kept signing without pausing. Again, pausing is important." (P1)

- "Yes, because hearing people sometimes talk too fast. It makes us lose information in our brain. If it is very important, they need to add pause because it helps to refresh the mind." (P3).

The results of the studies summarized in this section have established that for ASL-signers who view animations of ASL, the overall understandability and quality of the animation is influenced by the accuracy of the speed and timing parameters. These findings provide key motivation for the dissertation research to investigate this aspect of generating ASL animations.

2.8 Summary

This chapter has introduced various important concepts related to speed and timing in ASL signing and animation, with a goal of establishing key terminology to be used throughout this dissertation. Specifically, we have differentiated the concepts of speed and acceleration. We have also discussed how there has been significant prior research on human biomechanics and virtual human animation, yet there has not been sufficient research on how to specifically adjust speed and timing in ASL, to produce realistic ASL animations. As additional key vocabulary, we have presented a set of parameters that represent ASL speed. For instance, we considered how the speed of an ASL sign may depend upon the fundamental rate of the human signing (overall words per second), the base duration for the word (as the duration of the word had been defined in the system's lexicon), and the differential signing rate (how the performance of an individual word in a particular sentence may be faster or slower). In addition, when considering pauses during ASL signing, we discussed how there is a need to select where to insert pauses and to select what the duration of each pause should be. Finally, we presented some evidence from prior research that when DHH individuals view animations of ASL, they are sensitive to these speed and timing factors, which influence subjective impressions of the animation quality and viewer's comprehension of the information content.

Chapter 3

Prior Methods of ASL Generation and Synthesis

This chapter provide the reader with an overview of the state-of-art techniques used to produce sign-language animation, with a particular focus on how prior research has selected values of speed and timing for such animations.

Specifically, this chapter has been organized as follows:

- **Section 3.1 Related Work in Speech Synthesis** briefly summarizes some research in the field of speech synthesis which is analogous to some of the speed and timing prediction issues that are investigated in this dissertation in regard to ASL.

- **Section 3.2 Data-Driven Sign Language Animation** will address prior research into using data driven approaches for sign language (broadly, without a specific focus on speed or timing issues). This section will begin by discussing published research on various sign languages throughout the world, including prior work that has used data from motion-capture recordings as well as work that did not use such data. Then, the chapter will narrow its focus specifically to prior work on ASL (both prior work that did and did not use motion-capture data sources). Within the discussion of prior ASL motion-capture data, this chapter will

present various current corpora resources and discuss some limitations of existing datasets, which motivated our enhancements to an existing dataset in this work.

- **Section 3.3 "Speed and Timing in Sign Language Animations"** will narrow the focus to prior research on modeling and predicting speed or timing parameters to synthesize animations of sign language. This section will use the five-component model[5] of speed and timing outlined in Chapter 2 to organize its discussion of the literature.

- After generating the ASL animations, we need to know how to judge the quality of the generated animations, **Section 3.4** will discuss how prior researchers have evaluated the quality of generated animation systems.

3.1 Related Work in Speech Synthesis

There has been prior work in the field of speech synthesis that has examined the prediction of pauses (prosodic breaks) in speech. This is a necessary step in the pipeline of most speech synthesis systems; for instance, the Festival system uses a part-of-speech-based model for predicting such pauses [133]. Other work has built models of pause prediction specific to particular styles of text [114]. Some authors have examined the insertion of pauses in speech synthesis for low-resource languages [113]. After first building a part-of-speech tagger for the language using unsupervised techniques, these authors built a model to predict phrase boundaries (given the part-of-speech sequence), which they used to insert pauses. Inducing a part-of-speech tagger for ASL is impractical given the extremely small corpora (even unannotated) which are available, and due to linguistic aspects of the language that would make unsupervised part-of-speech tagging techniques challenging: ASL has little morphology that may indicate part-of-speech of words, and there are few function words that may indicate the parts-of-speech of surrounding words. For this reason, we investigate methods of predicting pause locations in ASL that do not depend on first creating a part-of-speech tagger and that require only the smallest number of annotation input.

[5] As a reminder to the reader, this five-component model included: Fundamental rate in animation systems (Subsection 2.4.1), base duration in animation systems (Subsection 2.4.2), differential signing rate in animation systems (Subsection 2.4.3), pause insertion in animation systems (Subsection 2.5.1), and pause duration in animation systems (Subsection 2.5.2).

3.2 Data-Driven Sign Language Animation

Many prior researchers who have attempted to automate the creation of sign-language animations have written specific rules for setting numerical values of speed and timing, to govern an animated character, in an attempt to produce natural movements. In this dissertation, we refer to this approach as *rule-base modeling* for ASL.

Many prior sign-language animation systems have encoded simple rules (details in Subsection 3.3.5) for planning speed and timing in animations, e.g. [17, 50, 132], where ASL professional animators synthesize a set of signs from a dictionary, and the programmers code some rules for blending adjacent signs together. In prior research (e.g. [2, 60, 63]), our lab had also performed some rule-based modeling for speed and timing of ASL animations, with the numerical parameters in the rules based on prior ASL linguistics literature.

In contrast, we will use the term *data-driven modeling* to refer to machine-learning-based predictive systems which are trained on datasets of many examples of human language, often referred to as corpora. The motivation for data-driven approaches is that, given the complexities of human language, it is generally difficult to use a rule-based approach to handle all cases that may arise, then data driven modeling has been used. In fact, many advances in computational linguistics in the past two decades have come from data-driven methods based on machine-learning models for this very reason.

Despite this trend, most prior work on sign language animation synthesis has been rule-based (rather than data driven) because there are few available video recordings that have been linguistically-annotated. For "low resource" languages, many data-driven methods may not work (without special adaptations) because we do not have enough data. Over the past decade, some small corpora for ASL have become available, and as these new resources emerge, it is becoming increasingly possible for sign-language animation researchers to attempt data-driven approaches [65]. Specifically, performance of a human signer can be recorded through video or a motion-capture; then, human experts transcribe and annotate this data by adding time-stamped linguistic information.

The remainder of this section will survey prior data-driven research on sign-language animation synthesis, beginning with work that has focused on other sign languages, and then focusing specifically on ASL. The discussion below will include prior research that has included machine-learning methods trained on annotated corpora that include motion-capture data from humans.

Some researchers have examined data-driven methods (that used video-based recordings of sign language, rather than motion-capture data from humans) for sign languages synthesis: Bungeroth

et al. [26] collected and annotated a corpus for German Sign Language (GSL) based on German television weather reports. The corpus annotations have both German sentences and GSL gloss which includes some important information like sentence boundary and part-of-speech tags. This work studied machine-translation and facial-expression issues, but not speed or timing. In later work, Morrissey and Way [106] investigated example-based machine translation approaches (using statistical machine translation techniques) for producing Deutsche sign language from English text, using a corpus, which they annotated with manual and non-manual features. Their corpus of Dutch Sign Language was annotated with time-aligned translation of English and Dutch. They generated word sequences for sign language, but not animation output nor any speed encoding, which would have required speed or timing information [106]. In both cases, Bungeroth et al. [26] and Morrissey and Way [106] made use of somewhat small video-based corpora on a narrow topic/domain, but neither had explicitly modeled speed and pausing of signs. In other work, Crasborn et al. [32], built a video recording corpus for Sign Language of the Netherlands (Nederlandse Gebarentaal: NGT), which included a wider variety of topics and sentence structures. This work annotated multi-sentence stories, and the authors provide open access online for the corpus data. In all these works, researchers in these studies did not generate actual animation output; instead, they produced video corpora for linguistic analysis or produced script representations of the sign-language message. As we will discuss below, generating animation and evaluating it among human ASL signers is important for rigorously evaluating research in this area.

Other researchers have made use of motion-capture data of humans performing sign language to investigate synthesis of animations, for example: Duarte et al. used a motion-capture approach to collect and build a dataset of French Sign Language (LSF) for the SignCom project [37, 45]. In that work, they synthesized novel sign language animations via reassembling elements of the recordings. The idea was to modify the grammatical structure to get understandable output; for example, to synthesize the sign "INVITE" in the sentence "I invite you," the authors used a recording of the verb "INVITE" but played it in reverse to produce a sign that moved in the correct direction through space for the sentence "You invite me." To synthesize sign language animations, the authors represented each sign as different channels; each channel holds partial information about the sign language animations, for example: channels of eye, arms, head, and etc. This prior work did not examine the issue of speed and timing during multi-sentence sign-language animations.

In other work, researchers have collected motion-capture recordings of individual signs. Cox et al. [31] built a motion-capture corpus of individual signs of British Sign Language (BSL). They

built and evaluated a system called "TESSA" for converting English speech to British Sign Language animations. The authors used motion capture approach to collect a recording and build a small corpus, consist of 11 BSL signs. Focusing on the domain of typical conversations at the customer-service desk of a post office, they used a few template-like phrases to build a limited set of sign language sentences [31]. Since their system filled words into templates which required unlimited number of templates (rather than synthesizing complete phrases), they did not address timing and pausing issues, which is the focus of our work. Campr et al. [29], built and annotated a motion-capture corpus of individual signs for Czech Sign Language. Their aim had been to use this dataset to create a sign language recognition system. Their corpus contains videos and whole-body 3D motion-capture data with facial expression and eye feature extraction. Their corpus consisted of *single-word* recordings only, which makes it ill-suited to learning any patterns related to sentence structures.

Some prior researchers have specifically focused on American Sign Language (which is the focus of this dissertation), including work that has used video-based corpora: For instance, Toro [136] collected video samples and designed animation algorithms for creating ASL animations. The focus of this work is creating ASL inflecting verbs using some human annotations.

There have also been several linguistics research projects that have collected some video corpora of ASL, with linguistic annotation, e.g. [2,88,110,136]. However, most prior corpora consist of single-word recordings, individual sentences, or pre-scripted messages. For learning the speed and timing patterns for ASL animation, it would be more useful to have multi-sentence, unscripted corpora because that is useful producing natural ASL animations; in addition, having direct recording of body motion using motion-capture equipment would make it easier to extract subtle timing details from the recording.

In prior research at our laboratory, our team had collected a motion-capture dataset of recordings of ASL from several ASL signers. This motion-capture corpus consists of video recordings of nine participants (native ASL signers) performing multi-sentence, unscripted passages in ASL. The corpus contains video and motion-capture data recordings of the handshape, hand movements, body position, and other details, and the data was subsequently annotated by a team of Deaf native ASL signers and linguists, who labeled the individual words, as well as some syntactic information such as sentence structure. For this dissertation, we used the first release of this corpus [100], which consists of 83 passages, performed by a total of 3 ASL signers, containing a total of 7,138 words. This corpus forms the basis of research described in later chapters of this document, and the en-

hancement of this corpus as part of this dissertation research (to make it suitable for research on ASL speed and timing) is discussed in Chapter 4.

3.3 Speed and Timing in Sign Language Animations

Further narrowing our focus on sign-language synthesis research that has considered issues of speed and timing, this section will specifically focus on a small number of prior research and commercial systems that have in some way attempted to model or predict speed and timing parameters for synthesized animations of sign language. For readers interested in a broader survey of prior research on sign-language animation synthesis more generally, please consult the survey in [65]. This section begins with a brief listing of the systems under discussion, showing the summary of this prior work in Table 3.1, followed by an analysis of how these systems may have addressed the five-component model of speed and timing which has been outlined above:

- **2008 Model of Huenerfauth [60, 63]**: The current state-of-the-art model for predicting these speed and timing parameters in ASL animation was presented by Huenerfauth at *ASSETS'08* [60]; he studied how to control an animated character to produce ASL with natural pauses and timing. Because we often compare our current system to this model, we shall refer to this with the short name "2008 Model." This model consisted of several rules for calculating various timing parameters, but it required substantial (and time-consuming) input information, namely a full syntactic parse tree for every ASL sentence. This model utilized linguistic findings from [52, 53, 55]. Huenerfauth designed two algorithms, for **Sign-Duration** and for **Pause-Insertion**, to calculate sign duration time and to calculate pause location and length; more details about these algorithms will be provided soon. It is good to mention that Huenerfauth conducted user studies [63] to evaluate his rule-based approach, and he demonstrated that inserting linguistically motivated pauses and linguistically altering sign-duration in the ASL animations, using his rules, increased signers' performance on a comprehension task.

- **Adamo-Villani and Wilbur [2]**: This was a rule-based system for generating sign speed and pauses, to add multiple prosodic elements into an animation of ASL being authored, based upon linguistic findings from [118, 143]. Their system predicted how to add a variety of prosodic enhancements to ASL animations, including: insertion of pauses and phrase-final

lengthening of sign duration. Villani and Wilbur's evaluation with users showed promising results from using this algorithmic approach to add prosodic features.

- **eSign [77]**: This project developed an animated signing avatar which could be used to convey information on European websites, e.g. government agencies; the system performed sequences of signs from a lexicon, based on a script designed by an author. This project has produced technologies for content developers to build sign databases using a symbolic notation, however, this approach do not model the complex aspects of sign language.

- **Ebling and Glauert [38]**: built a system for translating train announcements from German text to Swiss German Sign Language using the JASigning animation platform. The authors wrote a rule to insert a short pause after each item in lists, based on a suggestion from deaf users who viewed their system's animation output; however, their paper did not provide any general rule for when pauses should be inserted nor what their duration should be [39].

- **Segouat and Braffort [126]**: Created a French Sign Language corpus of motion-capture recordings and built an animation system that combined different elements of human motion to create novel sign language sentences. The annotation of this work were not made by native signers but the researcher themselves, which could result in the poor quality of annotation and any animation produced afterward. Their aim was to present information (like warning information and delay of trip) in French Sign Language in railway stations. Their system used data from their motion-capture corpus to generate novel sentences, with an algorithm for blending the motion in-between the hand positions before and after the sign. Segouat and Braffort used a rotoscopy technique (which is an animation technique in which motion data is produced by someone "tracing" on top of a video image) to study co-articulation (how the movements of the hands at the end of one sign are influenced by the beginning of the next) in sign language. Despite collecting a small corpus of LSF, they did not model speed or other timing issues.

- **Sign Smith Studio (SSS) [133]**: This was a commercially available product allowing users to create an animation of ASL, by assembling a sequence of words (from a lexicon provided) on a timeline and add other details. This tool provide ability to the user to make general manual adjustments on timing parameters.

Table 3.1: Related works on ASL animations

System	Which Sign Language?	Domain	Use motion capture?	Approach	Build animation system?	Speed and Timing	Evaluate animation's quality?
2008 Model of Huenerfauth [60, 63]	American Sign Language	Build ASL animation system	No	Rule-based	Yes	Algorithms	Yes
Adamo-Villani and Wilbur [2]	American Sign Language	Adding predictable prosodic markers to sentences	No	Rule-based	No	Algorithms	No
eSign [77]	British Sign Language	Add signing avatar to a domain specific website	No	Data-Driven	Yes	Examined the speed of human signers in videos performing	Yes
Ebling and Glauert [38]	Swiss German Sign Language	Translating announcements in a train station	No	Data-Driven	Yes	Rules to insert a short pause in the lists	Yes
Segouat and Braffort [126]	French Sign Language	Modeling coarticulation in animation avatar	Yes	Data-Driven	Yes	They did not model speed	No
Sign Smith Studio [133]	American Sign Language	Animation generation software	No	Manual	Yes	ASL expert manually select values for timing	Yes

Below, we discuss each of the five components of speed and timing, in regard to the systems listed above.

3.3.1 Fundamental Rate in Animation Systems

Sign Smith Studio [133] is a good example of an ASL animation scripting tool that enables a human user, who is knowledgeable of sign language, to create a message. This system did not have any automatic algorithms for predicting speed or timing factors, but it provided the human author of the message with a great degree of control, e.g. enabling users to adjust an individual speed multiplier for each word, modify the speed of the transitional movement between words, or add additional pause time (of any duration) after a word. This level of control came at the expense of effort from the author, who had to manually adjust all of these parameters, which could be difficult or time-consuming [133]. Notably, in regard to fundamental rate, the system provided the author with a "master speed control" slider that could be used to adjust the overall speed of the signer's movement. As discussed in a previous chapter, prior researchers have empirically investigated what fundamental rate of signing is preferred by DHH individuals watching animations of ASL. For instance, Huenerfauth in [60, 63] found that native ASL signers in an experimental study, with animations displayed at several different fundamental rates, preferred animations with a rate of 0.9 signs per second. Presumably, the slower rate preferred for animation (compared to that for human signers [52]) is due to the animation being more difficult to understand than a video of a fluent human signer. As the quality of ASL animations improves, DHH users may prefer faster animations - to more efficiently consume information. As discussed in "Chapter 6" in the user study we conducted with animations displayed to DHH participants, we used animations with a fundamental rate of 0.8 sign per second, after pauses were inserted.

3.3.2 Base Duration in Animation Systems

Most sign-language animation systems include a lexicon of individual signs that can be assembled into longer sequences. In such systems, each individual word entry in the lexicon will generally include information about the base duration of the sign (i.e. its default time duration). The source of this timing information has varied: In some cases, designers of individual words stored in the lexicon may have simply created animations without consulting any reference. Whereas, when creating the lexicon for the eSign system [77], the researchers chose the duration of each word by examining the

speed of human signers in videos [77]. Other researchers used recordings of human signers more directly: Segouat and Braffort [126] used rotoscopy to create a French Sign Language corpus, with the timing of the motion-data in their corpus based on the timing of the original video recording upon which each was based. This work was based on motion-capture data, as in our current work; here the authors re-combined elements of motion recordings to generate new animations and focus with the implementation on the co-articulation modeling (how adjacent signs interact) in ASL.

3.3.3 Differential Signing Rate in Animation Systems

As discussed in the Section 1.2 "Focus of This Dissertation," one of the aims of our study is to create a model that can predict the differential rate of speed of individual words in an ASL animation. In prior work, the rule-based system of Adamo-Villani and Wilbur [2] included an automatic rule for one form of differential rate modification: phrase-final lengthening of words. Specifically, they increased the length of the final sign in each phrase, following the prior findings of Wilbur [143]. The Sign-Duration algorithm of the 2008 Model of Huenerfauth [63] included similar phrase-final lengthening depended on whether specific signs had previously appeared in a performance and whether they are at the end of clauses, e.g. noun signs located at boundaries (sentence or clause); with signs extended 8% when immediately before a clause boundary, 12% when immediately before a sentence boundary and later appearance of verb signs is shortened in duration by a ratio of 12% (these percentages were based on linguistic findings in [52]). In addition, the 2008 Model modified the differential rate of signs (to become faster or slower) depending on whether a particular word had previously appeared in the passage (and whether its subsequent appearance was in the same syntactic position as its prior appearance). The values used for these rules were based on averages reported in the linguistics literature, not on any data-driven machine-learning method. As described in Section 1.2 "Focus of This Dissertation," we seek to create a machine-learning based model, which utilizes a variety of linguistic features.

3.3.4 Pause Insertion in Animation Systems

Few prior sign-language animation systems have attempted to predict where pauses should be inserted during the signing. Most simply ask the human author to indicate pause locations when using a scripting tool - or they simply insert a pause at every sentence boundary, which had been indicated by the human author. One exception is the train-station announcement system of Ebling

and Glauert [38], in which the authors inserted a hand-coded rule that added a brief pause in-between items in a list, but their rule only applied in this narrow context. The rule-based system of Adamo-Villani and Wilbur [2] automatically inserted pauses at specific locations between and within sentences, based on syntactic structural considerations described by Pfau [118]. Finally, the 2008 Model of Huenerfauth presented another algorithm called Pause-Insertion algorithm, which inserted pauses at inter-sign locations, based on the length of the current non-broken span of words and other syntactic factors, to insert pauses at 25% of word boundaries. This work was based on a prioritization scheme described in [63], which considered the entire syntactic parse tree structure of every sentence in the message. As described in our research method Chapter 5 "Selecting Data-Driven Models of ASL Speed & Timing," we investigate training a machine-learning classification model to determine, for every word boundary in a passage, whether a pause should be inserted at that location.

3.3.5 Pause Duration in Animation Systems

While human authors creating a sign-language animation have the ability to manually adjust the timing of pauses in some scripting systems, e.g. [133], few automatic algorithms for selecting pause duration in sign-language animations have been proposed. An exception is the 2008 Model of Huenerfauth [63], which chose the duration of the pauses that it inserted via an algorithm based on the sentences' syntactic structure and on a preference for inserting pauses mid-way between previously placed pauses, following the approach of [53, 55]. However, a limitation of this model is that it required the user to provide a complete syntactic parse tree of every sentence in the message, and the algorithm was based on rules and guidance in prior linguistic literature, which itself was based on researchers' observation of a small number of videos of ASL. We would instead prefer a model of ASL timing based on actual behavioral data and movements of native ASL signers performing fluent ASL passages.

3.4 Evaluating the Quality of Output ASL

So far, we addressed different methods used by researchers for generating ASL animations, but after generating an output ASL animation, we must select how to evaluate that animation, in particular

to determine whether the output is acceptable to users. This section discusses two methods of evaluation: dataset-validation model evaluation and user-based evaluation of animations.

3.4.1 Dataset-Based Validation Evaluation

In traditional evaluation of machine learning models, researchers train their model on a subset of the data and then evaluate whether the model can predict the pattern in the remaining data (in our case, where a human actually paused in an ASL recording and what is the signing speed for ASL animations). There are many approaches used to *split* the data in the model assessment and selection process for example, splitting the data to three subsets, typically a: training set, validation set, and test set. The training set consists of a portion of the data used to fit the model. The validation set contains samples of the original data used to provide an unbiased evaluation of each model during the model hyperparameters tuning process while fitting the training dataset. The test data is a portion of the data used to provide the confident and final unbiased evaluation of the selected or final model that has fit the training dataset; the test data is usually called unseen data because the selected model will never see these data till the prediction stage. This type of data can present a real world data that the model will see later. Figure 3.1 includes a simple illustration of this division of the data.

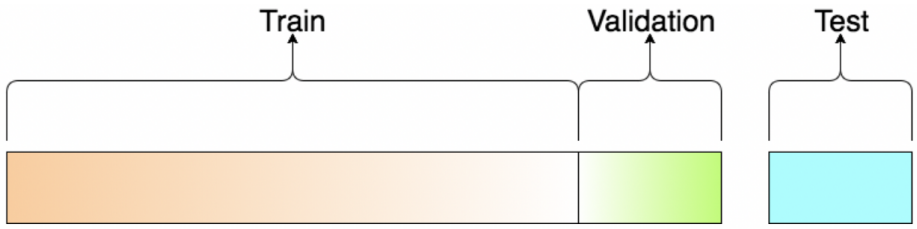


Figure 3.1: Data splitting

The training, validation, and test data will be used in an operation called cross-validation. Cross-validation is a technique used to estimate how the predicted model will generalize to future unseen data. There are many ways to perform cross-validation for example, “Leave-p-out cross-validation”, “Leave-one-out cross-validation”, “k-fold cross-validation”,... etc. When data is scarce, researchers often use “k-fold cross-validation” training and testing, in which the dataset is separated into a number of equally sized groups called folds (the commonly used cross-validation is 5-fold cross-

validation and 10-fold cross-validation [58]). The model is train on a subset of the data (excluding one-fold), and it is tested on this left-out fold. This process is repeated until all folds get a chance to be the left-out one. Accuracy is calculated by averaging across all folds [14, 92, 142]. After selecting the best model using the model-assessment and selection process summarized above, in this thesis, we are comparing our best model with the state-of-the-art rule-based "2008 Model" (explained in Section 5.7). We will refer to this approach of evaluation as dataset-based evaluation.

3.4.2 User-Based Evaluation of Animations

Since the goal of our research is to produce better animations of ASL for people who use ASL, it is very important to conduct an evaluation of the system with these users. Therefor, as a second level of evaluation, we will conduct experimental studies of animation quality with DHH users.

There are many factors that should be considered when conducting user-based evaluation study. In this section we will discuss the design protocol for the user study [75], the stimuli for the user study [62, 87], and the types of questions used in the user study [67, 85, 87].

The authors in [75] discussed the protocol and different requirements for identifying the level of ASL skill, demographic characteristics, and prior experiences of participants, when evaluating the usability and understandability of human-like animations. For instance, it is important to determine the age at which someone began to use ASL, as well as their early history in using the language during childhood, to determine whether someone is actually a native ASL signer. Such users were found to be the most discerning participants when evaluating animations of ASL [75].

Kacorri et al. investigated how to best engineer the stimuli and questions that could be used in a user-based evaluation study of linguistic facial expressions [87]. This work examines how changing a variety of user study parameters would impact the outcome of a user study. Regardless that this work focuses on facial expressions, the user study of this work illustrates important factors that should be considered during designing the user-based studies with DHH users. For example, the authors show that it is important to involve native ASL signers in the process of design study stimuli, and the appropriate language should be used for the study stimuli.

3.5 Conclusion

This chapter has presented a review of prior literature on data-driven research that investigated different aspects of sign languages, and then the chapter has discussed prior work that primary focused on American Sign Language. The chapter has discussed how researchers can make use of sign-language corpora, and how ASL animations generation researchers have addressed the five timing parameters in ASL. Finally, we presented two type of evaluations that we intend to use in this dissertation: dataset-based and user-based evaluation.

Chapter 4

Creating the ASL Speed and Timing Dataset

In this chapter we discuss the details of preparing our dataset used for speed and timing research for ASL, which was based on an existing motion-capture corpus of ASL that our laboratory had produced in prior work. Specifically, this chapter addresses the first contribution of this dissertation (which had been discussed in Section 1.3), as a reminder for the reader these contributions were:

Contribution 1: We have created a new American Sign Language Speed and Timing Dataset, which is an enhancement to our lab's pre-existing motion-capture corpus of ASL. As part of this work, we transferred our prior motion-capture corpus to a new linguistic annotation platform that has become standard among sign-language linguistic researchers, ELAN [130]. We have added layers of annotations and document our data preprocessing procedures which were necessary to make this resource useful for speed and timing research. We have also documented our feature engineering process to create input for machine-learning modeling, so that it is easier for future researchers to work with this new dataset.

This chapter begins by describing the existing motion-capture corpus that our lab had collected in prior work but had not previously used for any speed and timing research. Then, this chapter will

discuss how we enhanced and processed this corpus to create our ASL Speed and Timing Dataset, including the process of adding different layers of linguistic annotation and the data-processing workflow necessary to make this ASL dataset useful for ASL animation modeling.

4.1 Original Motion-Capture Corpus

During a five-year period from 2009 to 2013, our laboratory (the Linguistic and Assistive Technologies Lab, or LATLab) gathered video and motion-capture recordings from 9 native ASL signers, as part of an NSF-funded research project to create a linguistically annotated motion-capture corpus of ASL. In total, 246 unscripted multi-sentence single signer passages were recorded, and in 2013, the first sub-portion of this corpus (which had been cleaned and checked for quality) was released for research community [98, 100]. The first released sub-portion consist of 98 passages performed by three signers. Of the 3 signers in the first corpus release: all of them grew up with family member (father/mother) who was fluent in ASL, all of them considered ASL as their primary language at home, all used ASL at work, all had attended a university that used ASL as primary language of interaction, and one of the three participants was married to someone deaf/Deaf [99]. All of the three signers were young men aged between 22 to 33 years (average age: 25.7).

The nature of recorded stories in this corpus had been designed to be unscripted and fluent ASL: All of the interactions during the recording sessions were conducted in ASL, with a moderator who was also a native ASL signer, and the participant was given various prompts that were meant to elicit a few minutes of ASL signing monologue. The resulting passages that were recorded cover a variety of topics, including some signers discussing their own life, comparing between people or movies, sharing their opinion about a hypothetical situation, explaining a Wikipedia article they had read, explaining the story behind a set of photos, or sharing the plot of a book or movie [98]. Table 4.1[6] summarizes the types of prompt used to collect the recorded passages in this corpus.

Subsequent to collecting these recordings, the resulting videos were annotated by expert ASL signers (who is familiar with annotation tool), who added various types of linguistic annotations (explained below), using the SignStream ASL analysis software [109]. That software enabled the annotators to view the multiple camera views of each recording simultaneously and to produce a timeline with various parallel tracks of linguistic information, based on what had occurred in the

[6]Table reproduced from [99], and presents the English version of the prompt that were giving in ASL

Table 4.1: Types of used prompts.

Type of Prompt	**Description of This Prompting Strategy**
News Story	Please read this brief news article (about a funny or memorable occurrence) and recount the article.
Compare (people)	Compare two people you know: your parents, some friends, family members, etc.
Compare (not people)	Compare two things: e.g. Mac vs. PC, Democrats vs. Republicans, high school vs. college, Gallaudet University vs. NTID, travelling by plane vs. travelling by car, etc.
Photo Page	Look at this page of photos (of people who are in the news recently) and then explain what is going on with them.
Personal Narrative!	Please tell a story about an experience that you had personally.
Personal Intro/Info	Introduce yourself, describe some of your background, hobbies, family, education, etc.
Recount Movie Book	Recall a book you've read recently or a movie you saw, and then explain the story as you remember it.
Opinion / Explain Topic!	Please explain your opinion on this topic (given) or explain the concept as if you were teaching it to someone.
Wikipedia Article	Read a brief Wikipedia article on some topic and then explain/recount the information from the article.

video recording. These tiers of annotation included: the sign glosses (English labels representing which word was being performed); a part-of-speech tag for each gloss (e.g. noun, verb); syntactic structural elements that spanned multiple words (e.g. sentence, clause, noun phrase, verb phrase); and other non-manual elements such as linguistic face and head movements (e.g., yes-no questions, WH-word questions, topicalization, negation, conditionals, and rhetorical questions). The SignStream datafiles were in a proprietary file format, but the annotations could be exported into an ASCII plaintext format, with numerical references that refer to time (in milliseconds) during a video when each linguistic element began or ended.

In the original release of the ASL motion-capture corpus, a variety of file formats of data were shared, including:

- Autodesk MotionBuilder ".fbx" files: These files recorded the 3D movement of the skeleton of the human in each recorded performance, as determined by a set of sensors worn on the person's body. The MotionBuilder file format is a proprietary format of this Autodesk software, but it is commonly used in the human animation community. These files contain the original recordings and the virtual human character who represents the human signer's body proportions driven by the recordings.

- Bio Vision Hierarchical ".bvh" files: To make use of the original FBX files in analysis and processing tasks, those files had been converted into BVH files, which is a widely used animation file format. The BVH files are ASCII format files containing two types of information: (1) rows of numerical data that represent all of the skeleton joints on a frame-by-frame basis and (2) a hierarchy of body segments (skeleton bone) sizes and joints for the human.

- High-resolution ".mov" video files: Each signer's ASL performance was recorded from three angles using high-definition cameras: front view, side view, and face-close-up view.

- ASCII plaintext files containing exported linguistic annotation data from the original SignStream annotation files. (The original SignStream annotation files were retained internally at the laboratory, but only the plaintext extracted form of this data was shared.)

Table 4.3 contains additional technical details of the various files in the existing ASL corpus:

Table 4.3: Types of file in the Motion-Capture Corpus

File Type	**File Description**
MOV video format	The front, side, face views. Each video file corresponds to one "passage" that had been recorded in response to a prompt during the recording session. The three video files for each passage have been time-synchronized with each other, as well as with the BVH files and annotation files below.
FBX MotionBuilder format	Each FBX file contains motion-capture data corresponding from a human recording; however, each individual FBX file contains data for several passages which had been recorded in sequence during a recording session appointment at the lab. These data files were generated using Autodesk MotionBuilder 7.5, but could be imported into more recent versions of Autodesk MotionBuilder.
BVH skeleton files	The BVH files were exported from the FBX motion-capture files, and each individual passage that had been recorded (each in response to a single prompt) has been trimmed into an individual BVH file recording, which is time-synchronized with the videos above.
Annotation files	These ASCII files contain various forms of linguistic annotation extracted from the SignStream files, e.g. with one annotation file containing the gloss labels (with numerical values that correspond to the "frame numbers" of the video files above). Other annotation files contain data about noun phrases, verb phrases, etc.

4.1.1 What was Good About This Existing Corpus for ASL Speed and Timing Research?

As discussed in Chapter 3, while there had been several video-based corpora of ASL collected in prior research, e.g. [106, 136], our lab's prior ASL motion-capture corpus was the only publicly accessible corpus that contained actual body motion data. For investigating subtle details about speed and timing of human motion during ASL, having a dataset that captured subtle aspect of speed and acceleration is essential. In addition, this corpus contained linguistic annotation (with start-times and stop-times) for individual words and various syntactic phrases. As discussed in Chapter 2, various linguistic features of sentences may influence speed, timing, and pauses, and therefore having this information about each sentence would support our modeling of such relationships in the data. Having start and stop times for individual words and sentences will also enable us to identify the duration of inter-sign gaps, as judged by human annotators, which may provide a basis for investigating where prosodically-motivated pauses may be occurring during signing.

Other advantages of using this corpus for our research include characteristics of the signers and the recording process itself: The inclusion of native ASL signers, who were recorded in an ASL-based environment, suggests that the data is a good representation of fluent performances of ASL. In addition, the use of an unscripted prompt-driven recording approach means that the resulting passages are spontaneous and natural ASL signing. Further, since many aspects of speed and timing occur at a multi-sentence level, e.g. inter-sentential pauses, the fact that the videos are multi-sentence passages is useful for our future modeling research.

4.1.2 What Limitations did This Existing Corpus have for ASL Speed and Timing Research?

Although there are advantages to using this dataset, there are also several key limitations:

- The use of the proprietary SignStream annotation tool in that project led to a set of datafiles that are difficult to process and extract, and this format of linguistic annotation file is not commonly used among the research community. Instead, it would have been advantageous if the data had been annotated using a more common video-annotation software tool, such as ELAN [130], which is frequently used for sign-language research.

- Although the linguistic annotators in the original corpus creation process had labeled some syntactic information, e.g. noun phrase and verb phrase boundaries, a later analysis of the quality and completeness of that annotation at our laboratory revealed that most passages were missing this data, and the there were frequent errors in the annotation. To use this corpus for our speed and timing research, it was necessary to conduct a linguistic annotation project to re-label all of this syntactic information throughout the corpus.

- As discussed in Chapter 2, many phonological models of ASL are based on a Movement-Hold paradigm, in which features are represented at holds/keyframes during the performance, each with a time duration for how long the hand is stationary at this keyframe (potentially with a value of 0 if there the hand merely flows instantaneously through this keyframe), and moves/transitions flow in-between these holds/keyframes. None of this detailed phonological structure was captured within the original annotation of the lab's ASL motion-capture corpus, and therefore some post-processing of the motion-capture data (to identify potential "holds" in the recording) was necessary.

- Although the original dataset included annotation of when individual words began and ended, which thereby defines some inter-sign gaps in-between each word, there had not been any annotation of when a longer "pause" might have occurred at a subset of these inter-sign locations in the recording. Thus, some post-processing of the motion-capture data would be needed to identify longer-than-normal inter-sign gaps.

The remainder of this chapter will discuss activities as part of this dissertation research to enhance this original corpus to address some of these limitations.

4.2 Adding Additional Annotation and Building an ELAN Motion Capture Dataset

As shown in Subsection 4.1.2, the original motion capture corpus has some limitations, and we found it is a very challenging and time-consuming task to keep working on the original version of the motion capture corpus. In order to overcome these limitations, we produced a new version of the motion capture corpus that provides more flexibility and convenience for the ASL speed and timing research (and other researchers addressing other challenges in ASL). By adapting the

original motion capture corpus and building the new ELAN dataset we have discovered the following benefits:

- As we mentioned above, the original corpus had some missing annotations, therefore, we added new layers of annotation to the new ELAN dataset. For example, we have added: Clause layer of annotation, part-of-speech layer, and other layers of annotations. In this work we needed additional layers of linguistic annotations for two main reasons. First, based on some linguistic research on ASL, additional linguistic features are shown to improve the performance of the predictive models. Second, in the following stages of this work (specifically in the dataset-based evaluation) we need to compare our winning predictive models to the rule-based 2008 Model of Huenerfauth [60, 63] and this requires adding some layers of annotation to perform the comparison.

- So far in this work, we have used data from three ASL signers, but we have additional recordings ASL signers in our lab's motion capture corpus that have not yet been publicly released and have never been used in prior work. So, it is logical to process the additional data and make use of that data for the future modeling, given that some future research may need more data to use in deep learning modeling. So, in the new ELAN dataset we have confirmed the annotation of the three original signers of the original corpus, and we supervised the annotation of new data for one additional signer. We create new larger motion capture corpus that consists of four signers, who have the larger amount of data, using ELAN software.

- We documented the data processing steps and the approach we have used to engineer our set of features, which provide the ability to replicate this work. This may open the door for other research to use this dataset to investigate other aspects of ASL.

- The new ELAN-based version of the corpus is much easier to annotate because all the data (videos and annotation) are available in the same software window, which makes the annotation process more convenient for the annotator. The annotator has the ability to examine different views of ASL videos. Further, ELAN provides a flexible extraction functionality which makes the data extraction process much easier.

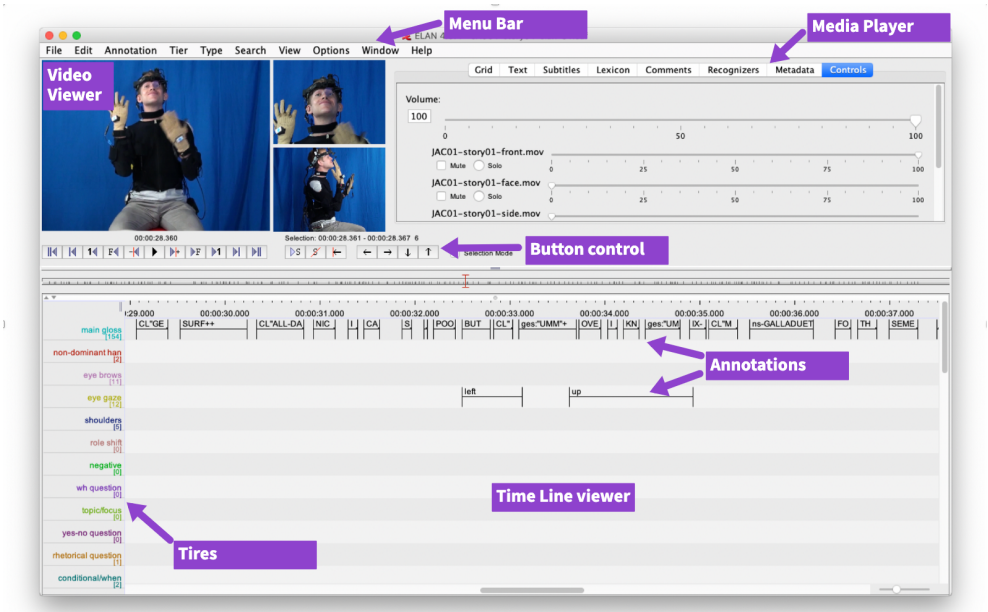


Figure 4.1: User-interface of the ELAN annotation tool

In this paragraph we are discussing the process of converting the original motion capture dataset to ELAN dataset. The original motion-capture corpus consists of a collection of stories. Each story had a SignStream annotation file that was associated with three source video files: The face view of the signer, the front view of the signer, and the side view of the signer (shown in Figure 4.1). We wrote code to extract the contents of the previous SignStream annotation files and generate the XML-based ELAN annotation files (.eaf) for each story. In the process of annotating a video, when the annotator edits the new .eaf files, ELAN will save the information inside the .eaf, and all changes are recorded in the annotation file; the source video file is left unchanged. This .eaf file links all of the original files for every single story in one screen window shown in Figure 4.1; and the annotator has the ability to view and control the videos using that window. The three videos (Figure 4.1) are represented on the top-left of the window. This includes the face view of the signer, the front view of the signer, and the side view of the signer. The top-right section of the window shown in Figure 4.1 contains different user-interface controls for the ELAN software. The annotation area is located on the bottom of the window. This is where the annotator has the ability to annotate different tiers (the term “tier” refers to one layer of linguistic annotation).

4.2.1 Dataset Annotation

After converting the original dataset to ELAN, we hired linguistic analysts (who were senior students training to become ASL interpreters who had completed courses on ASL linguistics for their degree program) to label the missing tiers from the original corpus according to a standard template that we previously used for annotating the original motion-capture corpus [99, 100]. A tutorial for using ELAN to label each of the linguistic tiers in our corpus was provided for the new ASL annotators, so the new ASL annotators would be able to more consistently label the corpus with the requested information. The annotators were asked to add the missing syntactic information like: clause boundaries, part-of-speech labels for words, noun phrase boundaries, and even other tiers which are not the primary focus of this work[7] . Specifically, we annotated 25 ASL specific tiers, 8 timing tiers, 10 tokens[8], and the English translation for the story.

While Appendix D presents the details of the annotation tutorial, it is briefly summarized here. We started by teaching the annotators how to install ELAN (on different operating systems), and we explained the various files used for annotation. Then we explained the ELAN graphical user interface GUI and the best settings of the software for annotation. Then we moved to detailed annotation guidelines for each tier, each of which focused on specific ASL language components. The tiers are visually divided into color-coded groups: blue, red, green, orange, and black. The different groups of tiers are: glosses, fingerspelling, visual non-manuals, abstract non-manuals, groupings, part of speech, tokens, and English translations. Understanding the meaning of these tiers is important to discussing the feature extraction in Section 4.3. The meaning of some of the annotation tiers:

- **Main Gloss**: The sign being performed; the annotator makes a gloss for each single sign.
- **Timing of Glosses**: The timing information includes when the gloss begins, and when it ends (when the hand begins to fall or move into the position of another sign). The times in which the hands are moving into position to make a sign are not included as part of the gloss. Similarly, we have identified the end point of the sign as occurring prior to movement of the hands out of the position for that sign in preparation for articulation of the following sign.

[7] The idea behind annotating the complete set of tiers is to make the new dataset have the complete annotations so it is easier to publish it later.

[8] This refers to the use of space around a signer for pronoun reference, and it was a focus of our lab's original ASL Motion Capture Corpus.

- **Noun Phrase**: Normally, a noun phrase consists of a main noun and some additional words around it, e.g. determiners or adjectives.
- **Verb Phrase**: The verb phrase has a verb as its head. Before the ASL main verb, there may appear a negative word (such as not or never) or an adverb phrase. If the verb takes a direct object, the direct object is part of the verb phrase. Adverbs, prepositional phrases, and adverbial clausesmay also appear after the main verb.
- **Part of Speech**: The annotator record the part of speech of each sign (noun, verb, prepositional phrase, and other elements). A detailed table for POS is available in Appendix D.
- **Dominant and non-dominant hand tiers**: We have used the dominant hand gloss tier for most of the information about manual signing. If for some reason the non-dominant hand is doing something unusual or different than what is happening on the dominant hand, then we add information to the non-dominant hand row. Most of the time, the non-dominant hand row is left blank.
- **Abstract Non-manuals**: We divided the abstract non-manuals to seven types including (role shift, negative, WH-question, topic/focus, yes-no question, rhetorical question, conditional/when). Table 4.5 present the description for the seven non-manual types.

4.3 Data Extracting and Pre-processing

In this section we are documenting how we processed motion capture data and how we extracted a large list of possible linguistic features, for our initial modeling research. This work represents the first attempt to use the original motion-capture corpus for speed and timing research; so, data preparation was anticipated to be a time-consuming aspect of this work. The large amount of time needed for this data processing was expected, as prior machine-learning research and development projects have often found data processing and cleansing consume a major portion of a project life cycle. We automated the data pre-processing task using some custom-written Python code to produce a “comma separate values” (CSV) output files. We achieve this automation by using a configuration file that stores different configuration parameters for the work, these parameters include: the paths for input and output directories, current signer code, the skeleton bone we are extracting, the required coordinates, and other configuration parameters.

Table 4.5: Abstract non-manual annotations

Non-manuals	**Descriptions**
Role Shift	Is the signer using the shoulder tilt to become a character in the story or to indicate one side of the signing space? The annotator should record when the signer becomes the character he is explaining, or imitating the facial expressions/quotations another person had said earlier.
Negation	Is the signer's head shaking left-to-right as in a negative manner? The annotator should record when the signer shakes his head when expressing negative opinion or expressing disagreement.
Wh-question	Is the signer making a WH-question facial expression? The annotator should watch when the signer raises his eyebrows and (often) tilts their head upward when asking a question and/or tilting their head a bit to the side.
Topics/focus	Is the signer raising his eyebrows for topicalized phrases at the start of sentence? The signer may raise his eyebrows at the beginning of sentence to present a new topic or emphasize the specific information as new topic. Sometimes, the signer move a noun phrase from its regular location in a sentence to the front of the sentence.
Yes-no question	Is the signer making a yes/no question facial expression? Yes-no question can include raised or lowered eyebrows to indicate a question such as "Are you all right?" or "Have we met before?"
Rhetorical Question	Is the signer asking a rhetorical question? The signer often uses "why" or "who" or "what" in the middle of sentence often replacing "because" from the English sentence but then quickly answering the question himself (e.g. ASL version: "I LOVE MOVIES, WHY?, THEY FUN WATCH" while the English version would state " I love movies because they are fun to watch.")
Conditional/When	Is the signer making the facial expression for a conditional "if" or "when" clause at the start of a sentence? Signer emphasize the "if" or "when" often with eyebrows or index finger pointing to time such as "1988, YEAR I BORN"

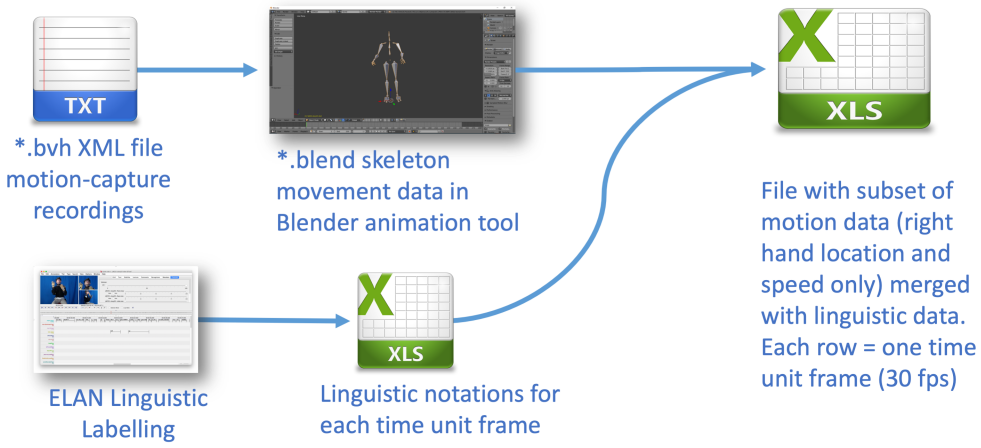


Figure 4.2: Data extraction and pre-processing

We needed to process and extract relevant information from motion-capture corpus, which contains: *motion-capture movement* data and *linguistic annotation*, as shown in Figure 4.2. The *motion-capture movement* data was available as .bvh files (Biovision Hierarchical Data) which is an XML file representing human joint angles from a movement recording. To support our processing of the movement data, we needed to convert each .bvh file into a .blend file, which is the standard input file format for the Blender animation software [30]. Then, within the Blender software, we extracted coordinates the X, Y, and Z coordinates (for the specified bone of the body) from *.blend skeleton animation. In this work since the signers in our corpus were all right-handed, we extracted the coordinates for the right hand wrist bone. Each bone in the Blender software is conceived of having a "head" and a "tail." In order to extract a location that corresponded to the right wrist joint of the signer, we extracted the "tail" position of the "right wristbone" of the human. The extracted information is used to calculate the speed[9] of signing.

The *linguistic annotation* was contained within the .eaf ELAN annotation file for each story. The linguistic annotation from these files was extracted and formatted. The outcome of this process is a

[9]For this dissertation research, it is important to document how the concept of speed is viewed, especially since ASL signs consist of periods of time when the hands are "holding" in a stationary position and periods of time when the hands are "moving" in-between these holds. For this dissertation work, we focus only on the sub-durations of time when the hands are moving, and we calculate the distance through space traveled in-between each sequential pair of frames. By summing these individual between-frame distances, we calculate a total distance traveled during a duration of movement. By dividing this distance by the time duration of this movement, a speed is calculated.

large CSV spreadsheet file for each story in the corpus. Then, we combined the different sub-stories for each signer to generate one large tabular file that consists of a series of organized columns and rows. This file has the following structure: Each row represented a single word in the corpus; and the columns of the file represented key linguistic features (which we intended to use for training our models). Table 4.7 presents the details of the columns in our generated CSV file. It is important to emphasize that not all of the CSV columns were subsequently used as features in our final models, as ablation analysis discussed in the next chapter led us to select a subset of these potential features. The selected features used in our model are therefore marked with value "Yes" in column "Used as a feature", and these features will discuss in detail in Chapter 5 when we explain the training of our models. The "Column Name" in Table 4.7 represents the name of the column in the generated CSV file, and the "Type" represents the data type of these features.

Within Table 4.7, in row #1, the Gloss is the sign being performed. Rows (#2-#5) list features, with each consisting of an ordinal value to mark the beginning, middle, and ending of the sentence, clause, noun phrase, and verb phrase consecutively. For these fields we used BMEWO encoding [22] which represent the boundaries for these phrases using end-of-entity (E_X) tokens, mid-entity tokens (M_X), and beginning-of-entity (B_X). For example, the noun phrase encoding is (B_N) for beginning of the noun phrase, (M_N) is the middle of the noun phrase, and (E_N) is for the end of the noun phrase. The next three rows (#6-#8) represent the length of each component of the sentence. Relative_Proximity (RP) is a numeric value representing how far the inter-sign gap appears from the midpoint of the current sentence. Complexity_Index (CI) represent the number of syntactic nodes that dominated this inter-sign gap. The formulas for Relative Proximity (RP) and Complexity Index (CI) are defined in [53, 60, 63]. While the definition in [53, 60, 63] depends upon knowing the full syntactic parse tree of the sentence, we do not make this assumption for our model. Thus, the Complexity Index used as a feature in our model training is based on the limited syntactic information available in our corpus (boundaries of sentences, clauses, verb phrases, or noun phrases). Table 4.9 presents our values for construction of the CI field based on the sentence syntactic structure locations.

Word_Duration and Next_Word_Duration contain the timing values for the current word and the following one. Word_Occurrence_Order is a counter that represents how many times this specific word has appeared in the story. Binary_Gloss_Occurrence_Order is another representation for the Word_Occurrence_Order field, but it represents a flag set as "True" if the word appeared more than once and "False" otherwise. We made used of Binary_Word_Occurrence_Order and Word_Occurrence_Order fields because we were inspired from linguistic research on the importance

Table 4.7: Set of column used to build features

Column #	**Column Name**	**Type**	**Used as a feature**
1	Gloss	Text	
2	Sentence Boundaries	Ordinal	Yes
3	Clause Boundaries	Binary	Yes
4	Noun Phrase Boundaries	Ordinal	Yes
5	Verb Phrase Boundaries	Ordinal	Yes
6	Sentence Length	Numerical	Yes
7	Noun_Phrase_Length	Numerical	Yes
8	Verb_Phrase_Length	Numerical	Yes
9	Relative_Proximity	Numerical	Yes
10	Complexity_Index	Numerical	Yes
11	Word_Duration	Numerical	Yes
12	Next_Word_Duration	Numerical	Yes
13	Word_Order_On_Sentence	Numerical	Yes
14	Reverse_Word_Order_On_Sentence	Numerical	Yes
15	Word_Occurrence_Order	Numerical	
16	Binary_Word_Occurrence_Order	Binary	
17	Independent_Clause	Text	
18	POS	Text	
19	Negative	Text	
20	whQuestion	Text	
21	yesNoQuestion	Text	
22	rhetoricalQuestion	Text	
23	topicFocus	Text	
24	conditionalWhen	Text	
25	Pausing	Binary	Yes
26	Pausing_Before_Gloss	Binary	Yes
27	Pause_Duration	Numerical	Yes
28	Pause_Duration_Before_Gloss	Numerical	

Table 4.9: Complexity Index as four level of representation

Sentence syntactic structure	Complexity Index (CI) value
Noun Phrase Boundaries	1
Verb Phrase Boundaries	2
Clause Boundaries	3
Sentence Boundaries	4

of these fields [52]; for example, Relative_Proximity (RP) and Complexity_Index (CI) had been used in prior rule-based projects for inserting pauses in ASL. Independent_Clause, POS, Negative, whQuestion, yesNoQuestion, rhetoricalQuestion, topicFocus, and conditionalWhen are representing different linguistic details of the ASL sentence, including several forms of non-manual information that represent, e.g., topic or conditional clauses.

In this research Pause is defined as a period of time where a human or avatar character slows or stops their motion for a greater than threshold (minimum amount of time) at a word boundary. Unfortunately, the original motion-capture corpus didn't have a labeling for the pauses so we needed to write a python code to calculate the pauses. The code extracts the movement from motion-capture corpus. To calculate Pauses, we identified moments in time in the data when the hand stopped moving for longer than one "frame" (1/30 of a second) near the end of a gloss (during a time span beginning two frames before the end of the gloss until the beginning of the next gloss). Then, we calculated the duration of this momentary "stop" of the hand movement. For each signer, across all of their "stops," we calculated the mean duration of such stops when they occurred at sentence boundaries. Finally, we calculated the Pause_Duration value, by subtracting the mean duration of the stops for that signer, from the duration of any specific stop.

4.4 Conclusion

This chapter presented an existing ASL motion-capture corpus collected in our laboratory's prior research, as well as limitations in that dataset that made it difficult to use for research on animation speed and timing. Next, the chapter has discussed the process of collecting, configurating, and the various types of files in that original motion-capture corpus, as well as how additional layers of linguistic annotations were added to the corpus using the ELAN tool by a team of ASL linguistic annotators. This work was followed by our processing of the sign language motion-capture corpus

and its linguistic annotations, to make the dataset useful for speed and timing machine learning modeling. The main focus of this chapter was to address Contribution 1 of this dissertation research: to create a dataset to support ASL speed and timing research. This chapter has documented the data extraction and processing steps needed to create this dataset, to provide documentation of this resource, so that it may be useful for future researchers.

EPILOGUE FOR PART I

In Part I, Chapter 2 introduced to the reader various definitions for speed and timing in ASL which represent the key concepts for the remaining parts of this work. We presented the sequential representations of ASL signs and animation. Then we illustrated the general pipeline for sign language generation. Focusing on sign language specifically, we presented five important speed and timing definitions which form the core components of the speed and timing concepts for this work. These five components are: fundamental signing rate, base duration, differential signing rate, pause insertion, and pause duration. Then we presented evidence from prior experimental work that had established that setting these speed and timing parameters correctly is important for users' comprehension and satisfaction, when synthesizing ASL animation.

In Chapter 3, we have discussed some prior work related to ASL animation generation. This chapter discussed prior work on speech synthesis, linguistics, and sign language technologies, including: how speech researchers address speed and timing challenges, linguistic research on speed and timing for ASL signing and spoken English, and data-driven sign language research. In addition, the chapter specifically focused on prior research related to our five components: fundamental rate in animation systems, base duration in animation systems, differential signing rate in animation systems, pause insertion in animation systems, and pause duration in animation systems. Finally, the chapter discussed some methodologies used in prior research, which inform our selection of options for different ways to evaluate our work in this dissertation.

In Chapter 4 we presented the process of collecting of the motion-capture corpus. Then we provided details about the process of extracting and cleaning of the first release of the corpus. In addition, we presented our approach of building the new ELAN version of our motion-capture corpus, how we added different layers linguistic annotations, and what is the best practice for preparing the data so that it can be used for modeling.

In summary, Part I of this dissertation has addressed the first contribution and established the groundwork for the subsequent contributions that have been listed in in Section 1.3 "Contributions of This Dissertation":

Contribution 1: We have created a new American Sign Language Speed and Timing Dataset, which is an enhancement to our lab's pre-existing motion-capture corpus of ASL. As part of this work, we transferred our prior motion-capture corpus to a new linguistic annotation platform that has become standard among sign-language linguistic researchers, ELAN [130]. We have added layers of annotations and document our data preprocessing procedures which were necessary to make this resource useful for speed and timing research. We have also documented our feature engineering process to create input for machine-learning modeling, so that it is easier for future researchers to work with this new dataset.

Part II: Modeling And Synthesis Of ASL Animation

PROLOGUE TO PART II

In Part II, we are focusing on creating ASL animations based on predictive models trained on human data. Chapter 5 presents how we engineered and selected the best subset of model features. Using the selected features, we will train three predictive models (ordered in a specific sequence) to predict three timing values in ASL: the insertion of pauses in ASL, the duration of the inserted pauses, and the signing rate within the ASL sentence. We will present the cross-validation results and the dataset-based evaluation of the modeling.

In Chapter 6, We will discuss a user-based evaluation, including conducting a study with DHH participants to learn their preferences about the animations generated with these models. We generate animations of ASL stories, and we conduct a user study where DHH participants saw different versions of animations of ASL stories, generated using our speed model, as compared to the baseline.

Part II will address the following contributions:

Contribution 2: We empirically determined which features were most influential in the speed and timing prediction models, e.g. via a feature-ablation analysis. Since our goal is to build a system that could convert a script that specifies an ASL message into an animation automatically, it is useful to identify a minimal set of information that the person writing the script must specify in order for our software to operate. We performed this analysis for each of the following three modeling tasks:

- **2.A:** Empirically determine the best subset of features needed to be used for building a predictive model for predication the **prosodic break (a pause)** after each word.

2.B: Empirically determine the best subset of features needed to be used for prediction the **time-duration of this break/pause**.

2.C: We empirically determine the best subset of features needed to be used for modeling the **variation of the speed for each particular word** in the message.

Contribution 3: We empirically determined whether a machine-learning modeling trained on a final subset of the linguistic features out-performs prior state-of-the-art rule-based approaches for the task of predicting the timing parameters for ASL multi-sentence passages. Specifically, in a cross-validation analysis of held-out data, we automatically identified the following three speed and timing values for each individual word in a message:

3.A: Is there a**prosodic break (a pause)** after this specific word? ASL signers will naturally pause at various locations during a message, typically more frequently at structural boundaries, e.g. as discussed in [118].

3.B: If so, what is the **time-duration of this break/pause**? ASL signers are also more likely to use longer pauses at more important structural boundaries [53].

3.C: Given the overall signing rate that we seek to produce, what is the **variation of this speed (slightly faster, slightly slower) for each particular word** in the message? ASL signers will generally slow down at the end of sentences, or change their signing speed for individual words, for a variety of reasons [52, 143].

Contribution 4: Empirically determine whether Deaf ASL signers prefer animations of multi-sentence ASL passages in which timing values are determined by these new models or by the previous state-of-the-art rule-based technique.

Chapter 5

Selecting Data-Driven Models of ASL Speed and Timing[10]

This Chapter discusses in detail the process of training several machine-learning predictive models for predicting speed and timing in ASL. We will start with explaining the feature engineering process. After selecting the best sub set of features we explain the process of selecting a robust model. The primary goals of this chapter are to investigate the second and third contributions of this dissertation research:

Contribution 2: We empirically determined which features were most influential in the speed and timing prediction models, e.g. via a feature-ablation analysis. Since our goal is to build a system that could convert a script that specifies an ASL message into an animation automatically, it is useful to identify a minimal set of information that

[10]The information in this chapter is based on several projects that include collaboration with other researchers in the CAIR lab at RIT (Larwan Berke, Sushant Kafle, Peter Yeung) supervised by my advisor (Dr. Matt Huenerfauth). The details of pause insertion modeling were published in our work [12], for which I was first author. Our modeling approach of the pause insertion, differential rate, and pause duration was presented at our paper at the ASSETS'18 conference [7], which received the best paper award and for which I was also first author. Furthermore, I have received valuable and important feedback when I presented the work from this chapter at the ASSETS'19 Doctoral Consortium [6].

the person writing the script must specify in order for our software to operate. We will perform this analysis for each of the following three modeling tasks:

2.A: Empirically determine the best subset of features needed to be used for building a predictive model for predication the **prosodic break (a pause)** after each word.

2.B: Empirically determine the best subset of features needed to be used for prediction the **time-duration of this break/pause**.

2.C: We empirically determined the best subset of features needed to be used for modeling the **variation of the speed for each particular word** in the message.

We need to identify a minimal set of features that the ASL human author, who is creating the script for the ASL message, should provide for our software to operate. If we are able to create a model that performs well with fewer features (i.e. using minimum human manual effort), then it would not be necessary for the human author to provide information for all other features we examined. Then, after selecting the best subset of features for modeling, we will make use of these features in the modeling step which is the focus of contribution three.

Contribution 3: We empirically determined whether a machine-learning modeling trained on a final subset of the linguistic features out-performs prior state-of-the-art rule-based approaches for the task of predicting the timing parameters for ASL multi-sentence passages. Specifically, in a cross-validation analysis of held-out data, we automatically identified the following three speed and timing values for each individual word in a message:

3.A: Is there a**prosodic break (a pause)** after this specific word? ASL signers will naturally pause at various locations during a message, typically more frequently at structural boundaries, e.g. as discussed in [118].

3.B: If so, what is the **time-duration of this break/pause**? ASL signers are also more likely to use longer pauses at more important structural boundaries [53].

3.C: Given the overall signing rate that we seek to produce, what is the **variation of this speed (slightly faster, slightly slower) for each particular word** in the message? ASL signers will generally slow down at the end of sentences, or change their signing speed for individual words, for a variety of reasons [52, 143].

After selecting the most accurate model, this chapter will explain our procedure for dataset-based evaluations by comparing the performance of our predictive models to the state-of-the-are 2008 Model [60, 63]. In order to conduct this comparison between our new models and this older model, which had required more linguistic features in order to operate, we had to first add more linguistic annotation to out corpus, to provide a complete syntactic parse for each sentence for a subset of our training corpus [60, 63]. In general, throughout our evaluation we investigated the following possibilities:

- In a cross-validation study (where models are trained and tested on various partitions of a dataset), does our model of where human ASL signers insert pauses in their ASL signing have higher accuracy than baseline models (to out-perform a baseline model that inserts pauses at the end of the sentences only or the prior state-of-the-art 2008 Model of Huenerfauth [54,56], which was a rule-based [60, 63] approach?)

- Given the predicted locations of the pauses, in our cross-validation study, will our model of the time-duration of pauses outperform a baseline (uniform duration) or the 2008 Model of Huenerfauth [60, 63]?

- In a cross-validation study, will our model of differential signing rate outperform a baseline (uniform speed) or the rule-based 2008 Model of Huenerfauth [60, 63]?

This chapter is organized as follows, Section 5.1 will discuss the feature selection process. Section 5.2 will present the logical sequence for building the models. Section 5.3 will address some assumptions that should be considered before modeling. Section 5.4, Section 5.5, and Section 5.6 will discuss the design, feature selection, and cross-validation evaluation of the three models (pause insertion, differential rate, pause duration). Finally, Section 5.7 will present a dataset-based evaluation of the models.

5.1 Feature Engineering

We had to invent linguistic features that are useful for modeling some linguistic phenomena and write code to extract these features from raw corpus data. The goal of the feature engineering step is to support our identification of effective models that are based on a minimal set of features, as to require as little input information as possible to the models. This would reduce the manual human effort and dedicated time which is very costly, while maintaining a robust model. As shown in Section 4.3, to create a dataset for our research, we needed to process the original motion-capture corpus. Specifically, we generated a CSV spreadsheet file for each story in the corpus, with the following structure: Each row represented a single word in the corpus; the columns of the file represented possible linguistic features. However, this file contains a large number of possible features, and we need to select a subset of these features which we intended to use for training our three timing models: pausing after words, time duration of the pause, differential signing speed. The outcome of our feature selection process is a set of the possible features that will be used for modeling, as shown in Table 5.1. All numerical features shown in the table were normalized to the scale (0-1). As shown in Table 5.1, the set of "predictor features" included information about whether this word was adjacent to a syntactic phrase boundary, the length of the current phrase or sentence in which the word occurs, how far this word is from the nearest pause in the signing, and some numerical measure of how major the syntactic boundary is that immediately follows this word. Also, in the feature engineering process we made use of other linguistic properties referred to as Relative Proximity (RP) and Complexity Index (CI) that had been used in the 2008 Model of Huenerfauth [60, 63]; those features had been inspired by the linguistic analysis methods of predicting speed and timing used in [60, 63]. However, our CI calculation differs slightly since we have only a partial parse tree from our annotators sentence, clause, verb phrase, and noun phrase spans only (as explained in Section 4.3). Table 5.1 also indicates which of our three models made use of each feature (this topic will be discussed in Section 5.2 and Section 5.3). The **Feature Type** column indicates whether the feature is numerical (Num.) or categorical (displaying the values). This chapter is presented here so that it may serve as a reference for the reader, as additional details about our investigation are discussed throughout this chapter. At this time, some of the columns or details in this table have not been fully discussed, but those details will emerge in our discussion of our modeling work throughout this chapter.

Table 5.1: List of predictor features used in this study, with a checkmark indicating if that feature used in each of our three models

Features Used in Each Model	**Feature Type**	**Pause Insertion**	**Differential Rate**	**Pause Duration**
#1-4: Is the gap after this word on the boundary of a sentence, clause, noun phrase, or verb phrase?	Yes / No	✓	✓	✓
#5: Relative Proximity (RP): How close is the gap after this word to the midpoint between the two nearest pauses?	Numerical	✓	✓	✓
#6: Complexity Index (CI): Value indicating the syntactic importance of this gap (ranging 1-4) with value of 4 at sentence boundaries.	Numerical	✓	✓	✓
#7-9: How many words are in the current sentence, noun phrase, and verb phrase (if applicable)?	Numerical	✓	✓	✓
#10-11: Is there a pause immediately before or immediately after this word? (This is output of the "Pause Insertion" model.)	Yes / No	X	✓	✓
#12-13: How far is this word from the beginning of the current sentence? From the end?	Numerical	X	✓	✓
#14-15: What is the differential rate for the current word and the following one? (This is the output of the "Differential Rate" model.)	Numerical	X	X	✓

5.2 Models Overview

We implement three machine-learning models to address various aspects of ASL speed and timing. Our models were cascaded, that is, an ordering was established among them such that the output of prior models could be used as an input feature to a subsequent model:

1. **Pause Insertion**: The first model was a classification model to determine if a pause should be inserted after the current word.

2. **Differential Rate**: The second model was a regression model to predict the change in signing rate within the ASL sentence. As shown in Table 5.1, this model used four more features than the Pause Insertion model (e.g. how close this word was to the end of the sentence). In addition, the Differential Rate model used the output of the Pause Insertion model as one of its input "predictor" features (to consider whether a pause had been inserted before or after the current word, which might suggest it would be performed more slowly, as the signer anticipated or resumed from a pause).

3. **Pause Duration**: The third model was a regression model to predict the length of each pause. This Pause Duration model logically depends on the results of Pause Insertion, since pause durations need only be calculated where pauses will occur. In addition, we utilized the output of the Differential Rate model as one of the input features to this model; specifically, we anticipate that pauses occurring near words with longer duration will themselves be longer in duration.

5.3 Important Assumptions Before Modeling

There are some important assumptions that we would like to present before moving to the details of model training.

5.3.1 Assumptions Used to Estimate Differential Speed

As discussed in Chapter 4, there are challenges in making use of ASL recordings to determine differential rate, since the observed timing of a word in a recording is also based on the signer's fundamental rate and the word's base duration. While we can estimate the fundamental rate of

the signer across a large sample of recordings, it is more difficult to estimate the base duration of each word. Without having hundreds of recordings of each word, as performed by a variety of signers, it is difficult to estimate what the "normal" base duration of each word is in ASL. For this reason, some researchers will examine speed or acceleration curves for hand movements during signing, e.g. [36], in lieu of considering the final duration of a word in a recording, when calculating differential rate. As suggested in our earlier discussion of this issue in Subsection 2.4.3, in our work, we have estimated the differential rate for each word using the following procedure: We focus on the movement of the signer's dominant hand (right hand of a right-handed person) only, and we omit any time frames during the word when the hand is stationary (since some signs contain periods of time when the hands make contact with the body or remain in place for short period of time). Next, we calculate the average velocity of the hand during these remaining time frames. Finally, we divide by the fundamental signing rate for this signer, as calculated across all recordings of this person in the corpus. In this way, if a word is performed more quickly in some context, we expect a differential rate greater than 1, and vice versa.

5.3.2 Assumptions Used to Estimate Pause Insertion & Duration

Our corpus did not include human judgements about where pauses were occurring nor what their duration was; thus, we also needed to estimate these values by processing the movement data in our recordings. As suggested by our earlier discussion in Section 4.3, a challenge is that many ASL signs will contain brief moments of time that are often referred to as "holds," in which the hand is momentarily stationary - and often these occur at the end of a sign. Such brief holds that are part of a sign performance are not generally considered to be a pause (i.e. they are part of the performance of the sign itself lexically, rather than being syntactically motivated prosodic phenomena). Thus, we needed to "filter" out these short holds at the end of words, to identify the slightly longer moments when the hands are stationary that reflect true "pauses." Taking guidance from Grosjean et al. [52] who observed that pauses occur at 25% of inter-sign locations, we extracted the end-of-sign stationary-hand durations for all words in the corpus, ranked these values, and decided that the top quartile (following the ratio published in [52]) of these values were "pauses." Furthermore, we determined that the "duration" of each of these pauses would be the amount of stationary-hand duration that was in excess of the duration value threshold that defined the top quartile.

5.4 Pause Insertion Modeling

The next section will cover three main objectives. The first is presenting the design of the Pause Inseration model, the second is selecting the best subset of features for modeling, and the last is the cross-validation result when evaluating the models from the first two objectives. This section will address **Contribution 2.A** and **Contribution 3.A**.

5.4.1 Design of Pause Insertion Model

For pause insertion model, we organized our dataset so that the first column is a "target" label that indicates whether this gap location in the corpus was where the human performed a "pause." The remaining columns contain properties about this gap location (e.g., is this a boundary between two sentences) that may be relevant to predicting pauses; we refer to these as "predictor" features. We wrote code to calculate sentence, clause, verb phrase, and noun phrase boundaries and lengths, along with other syntactic complexity features mentioned in Table 5.1.

Since our goal was to fit and test a model to predict pause locations in ASL animation and our target variable had values of ("there is a pause here" or "there is not a pause here"), we considered a traditional supervised classification approach to make an individualized prediction for the gap following each word in a sentence. Since we had both categorical and numerical predictor features (see the "Type" column in Table 5.1), we chose to investigate and compare several machine-learning algorithms that support mixed features, including: decision trees, support vector machines (SVM). In particular, we noted that prior work on pause prediction for English (Sarkar and Rao, 2015) or other modeling for ASL (Shibata et al., 2016) had successfully used decision-tree-based learning methods.

Aside from making independent predictions of the target variable ("pause" or "no pause") for each inter-sign gap location, we also investigated if there were dependencies between the values at subsequent gap locations. Specifically, we considered making predictions based on a +/-1 context window (i.e. the predictor features of the inter-sign gap immediately preceding and following the current inter-sign gap), thereby treating the problem as a sequence-tagging problem. For this purpose, we trained a Linear-Chain Conditional Random Field (CRF) model [21, 60], using [116]. CRF operates on the context-features and looks for the most optimal path through all possible target sequences for a sequence of words in a sentence.

5.4.2 Features used for Pause Insertion Model

We used the predictor screening tool form JMP Pro software [78] to select the optimal subset of features to use when building our model. Predictor screening mechanism uses bootstrap forest partitioning to evaluate the contribution of predictors towards the response; this mechanism gives us an initial set of possible features to be used for modeling. For example, we performed predictor screening to investigate the most important features to be used for pause prediction modeling; the report for these features is shown in Figure 5.1. The "Predictor" is the name of the possible features under investigation, and the "Contributions" column shows the contribution of each predictor feature to the bootstrap forest model. The rank and the bars in Figure 5.1 are provided for convenience as a simplified visualization of the results. Figure 5.1 indicates that the predictors with the highest contributions are likely to be important in predicting pausing. Therefore, we decide to eliminate the features that had less than 1% of contribution towards the model. A similar approach has been used to get a basic idea about the important features to be used for building other predictive models. Therefore, Subsection 5.5.2 and Subsection 5.6.2 will be much shorter than this section, but the same concept used in this section applied for both Subsection 5.5.2 and Subsection 5.6.2. Table 5.1 lists the various predictor features used in each model.

Specifically, for the features used for Pause Insertion model, the Pause Insertion model used the first nine features in the Table 5.1, including: whether the current location was phrase boundary, the syntactic importance of the boundary, and the proximity to nearby pauses.

	Pausing	
Predictor	**Contribution**	**Rank**
Sentence_Boundaries	27.4911	1
Complexity_Index	23.0773	2
Clause_Boundaries	12.2934	3
Word_Duration	7.9164	4
Next_Word_Duration	5.0951	5
Sentence_Length	4.7660	6
Relative_Proximity	4.3261	7
Verb_Phrase_Boundaries	2.4818	8
Word_Order_On_Sentence	2.0993	9
Word_Occurrence_Order	1.4227	10
Verb_Phrase_Length	1.2413	11
Binary_Word_Occurrence_Order	0.6949	12
Sentence_type	0.3501	13
Noun_Phrase_Boundaries	0.2782	14
Noun_Phrase_Length	0.2443	15
topicFocus	0.1200	16
conditionalWhen	0.0250	17
Negative	0.0071	18
rhetoricalQuestion	0.000868	19
independentClause	0.0000	20
POS	0.0000	21
whQuestion	0.0000	22
yesNoQuestion	0.0000	23

Figure 5.1: Predictor screening report, with a red dotted line show which features were chosen for the modeling

5.4.3 Pause Insertion Cross-Validation Model Evaluation

For the classifiers described in Subsection 5.4.1, we implemented a 5-fold cross-validation procedure, dividing our data into 80% training set and 20% testing set at each evaluation fold. We calculated the average accuracy and F-score across the 5 folds. We evaluated our model and compare our result with some baselines. Table 5.3 shows the accuracy and F-score for the model. For comparison, we also presented results for the two proposed baselines. Our baselines are defined as follows:

Table 5.3: Pause Prediction Model Results. The Decision Tree and SVM classifiers were implemented in MATLAB using the Classifier Learner Package, while the Linear-Chain CRF classifier was implemented using the sklearn-crfsuite package in Python. Parameter setting for the models are share in the footnote

Classifier	**Accuracy (F1-Score)**
Linear-Chain CRF[11]	**0.80**
Decision Tree[12]	0.76
SVM (Linear)[13]	0.76
Baseline 1	0.77
Baseline 2	0.64

- **Baseline 1**: This baseline inserts a pause at the end of every sentence (and nowhere else). The rationale for this baseline is that if a human animator were to create an animation and manually chose to insert some pauses, the animator may be likely to put them at all of the sentence boundaries, as a simple approach.

- **Baseline 2**: This baseline inserts a pause randomly at 25% of paragraph locations. To account for possible bias in evaluation, due to randomness, we ran it ten times (Table 5.3 presents the average).

As shown in Table 5.3, The linear-chain CRF model beat the proposed baseline, with an accuracy of 80% and with F-score slightly exceeding the baseline.

5.5 Differential Rate Modeling

In this section, we are investigating **Contribution 2.C** and **Contribution 3.C** which correspond to predicting the differential signing rate during ASL sentences, and selecting the best subset of features for differential signing rate modeling.

[11] **Function:** CRF.
Parameters: algorithm: l2sgd, c2: 0.0869, max_iterations: 100, all_possible_transitition: True
[12] **Function:** fitctree.
Parameters: SplitCriterion: gdi, MaxNumSplits: 100, Surrogate: off.
[13] **Function:** fitcsvm.
Parameters: KernelFunction: linear, PolynomialOrder: [], KernelScale: auto, BoxConstraint: 1.

5.5.1 Design of Differential Rate Model

Differential Signing Rate was modeled using regression, to predict a value that represents a multiplier to modify the Base Duration of a sign. During the initial training phase of selecting the best model we compared the performance of different algorithms including: Ada Boost (ABR), Gradient Boosting (GBR), Random Forest (RFR), and Extra Trees Regressors (ETR) [21, 124].

5.5.2 Feature Used for Differential Rate Model

We adopt the same principle used in Subsection 5.4.2 to create useful features for the differential rate model. We tested a set of features that can be used for modeling the differential rate, and we made use of the basic features used for pause insertion modeling. i.e. the first nine features of Table 5.1, including: the sentence boundary, clause boundary, noun phrase boundary, verb phrase boundary, the relative proximity, the complexity index, the location of this word inside the sentence/noun phrase/verb phrase, and how far is this word from the beginning of the current sentence. Since the outcome of the pause insertion model is the parameter "Pausing," it is now eligible to use as input. In addition, since we envision our model to process a script of an ASL sentence from left-to-right, we can not only consider whether there had been a pause immediately after the current word, but we can also consider whether there had been a pause immediately prior to the current word. These two additional pausing-related input features as shown as features 10 and 11 of Table 5.1.

5.5.3 Differential Rate Cross-Validation Model Evaluation

In a similar approach used for modeling the Pause Insertion, we implemented a 5-fold cross-validation procedure, by dividing our data into 80% training set and 20% testing set. We calculated the average "Root Mean Squared Error" (RMSE) across the 5 folds. To select the best working parameters for each of our models, we performed a grid-search to optimize the model performance. For the Gradient Boosting Regressor (GBR), we used GridSearchCV to exhaustively search the parameter space. We investigated the range of (50 to 400) for the estimators (n_estimators) parameter, and we found that a value of 50 led to the best result on our training dataset, while avoiding overfitting. In the Differential Signing Rate model, our baseline is the average signing rate for all signs in the corpus (i.e. predicting a uniform signing rate for all signs, specifically a "multiplier" of 1.0). Table 5.4 shows the performance as compared to this simple baseline. Our model outperformed the baseline;

a lower value for RMSE error is a better result. As shown in Table 5.4 GBR model had the best performance.

Table 5.4: Differential Signing Rate prediction model results

Regression Model	**RMSE**
ABR	0.56
ETR	0.55
GBR	**0.47**
RFR	0.53
Baseline	0.50

5.6 Pause Duration Modeling

In this section, we are investigating **Contribution 2.B** and **Contribution 3.B** which correspond to modeling the time duration of these pauses and identifying the best subset of features for that modeling.

5.6.1 Design of Pause Duration Model

Pause Duration was also modeled using regression, in this case, to predict a value for the time duration of the pause between two signs. We trained different ML classifiers such as: ABR, ETR, GBR, and RFR.

5.6.2 Feature Engineering for Pause Duration modeling

We made use of the same principle used in Subsection 5.4.2 and Subsection 5.5.2 to select the subset of useful features for the pause duration model. We made use of the first eleven features in Table 5.1: the sentence boundary, the clause boundary, the noun phrase boundary, the verb phrase boundary, the relative proximity, the complexity index, the location of this word inside the sentence/noun phrase/verb phrase, how far is this word from the beginning of the current sentence, and the existence of pausing before and after this word. As discussed in section 5.2, since the models are cascaded, and since the output of our Differential Rate model would already be known at this point we decided to use this value as an input feature for pause duration model. The differential-rate

features for the word preceding the boundary and for the word following the boundary are appear as features 14 and 15 of Table 5.1.

5.6.3 Pause Duration Cross-Validation Model

Table 5.5 shows the performance of the Pause Duration regression model, as compared to the proposed baseline (predicting uniform pause duration, specifically the average duration of all pauses in the corpus). GBR has the lowest error in Table 5.5 meaning that GBR outperformed the baseline with RMSE equal to 2.9, which is almost half of the error than the proposed baseline. In regard to parameter tuning, our grid search of the parameter space found that n_estimators=100 had the best performance on our training dataset.

Table 5.5: Pause Duration prediction model results

Regression Model	**RMSE**
ABR	3.54
ETR	3.2
GBR	**2.9**
RFR	3.0
Baseline	4.47

5.7 Dataset-Based Comparison to State-of-the-Art Model

So far, we showed that our model out-performed a baseline model (inserting pauses at sentence boundaries only, with uniform pause length and uniform sign speed); however, a better test of this new model would be to compare it against the current state-of-the-art model for speed and pausing in ASL, the 2008 Model of Huenerfauth [60, 63]. Aside from comparing model accuracy, there are other points of comparison between these models:

- The 2008 Model [60, 63] is a rule-based approach based on findings of prior linguistics research [52, 53, 55] that considered a relatively small number of ASL videos. Models based on larger datasets of human signing performance may be more accurate. Furthermore, the 2008 Model requires a full syntactic parse of every sentence as an input for its algorithms; this may be time-consuming for the human author of the ASL message to provide.

- Our new model, which we refer to as "ASL-Speed" is a collection of machine-learning models trained on human behavioral data, from a motion-capture corpus of ASL. ASL-Speed requires a smaller set of features (Table 5.1) as input from an ASL-knowledgeable human user, which uses less input data than required by the Huenerfauth [63] model, which needed a full syntactic parse tree.

Unfortunately, given the need for a full syntactic parse of all sentences in order to run the 2008 Model, running a test on the entire corpus [100] was impractical, since the original annotation of that corpus did not include such annotation. The corpus contained 83 multi-sentence passages, from a total of three signers. For this comparative analysis between ASL-Speed and the 2008 Model, we had to use a subset of the corpus. We selected three passages from each signer (selecting the three with length closest to the median for each person); there were 958 words total in the 9 passages. Next, an ASL linguist annotated each passage with a full syntactic parse tree. On this small testing set, we ran our new ASL-Speed model and the 2008 Model [63]. These passages were excluded from our models' training set. Appendix A provides more details about the selected stories for the 2008 Model and ASL-Speed comparison. Figure 5.2 shows the prediction accuracy for the new ASL-Speed model and 2008 Model for the Pause-Insertion task. The new ASL-Speed model had higher accuracy than the 2008 Model. Given that the 2008 Model had been the previous state-of-the-art method of predicting this information for ASL animations, this is an important indicator of the quality of our new model.

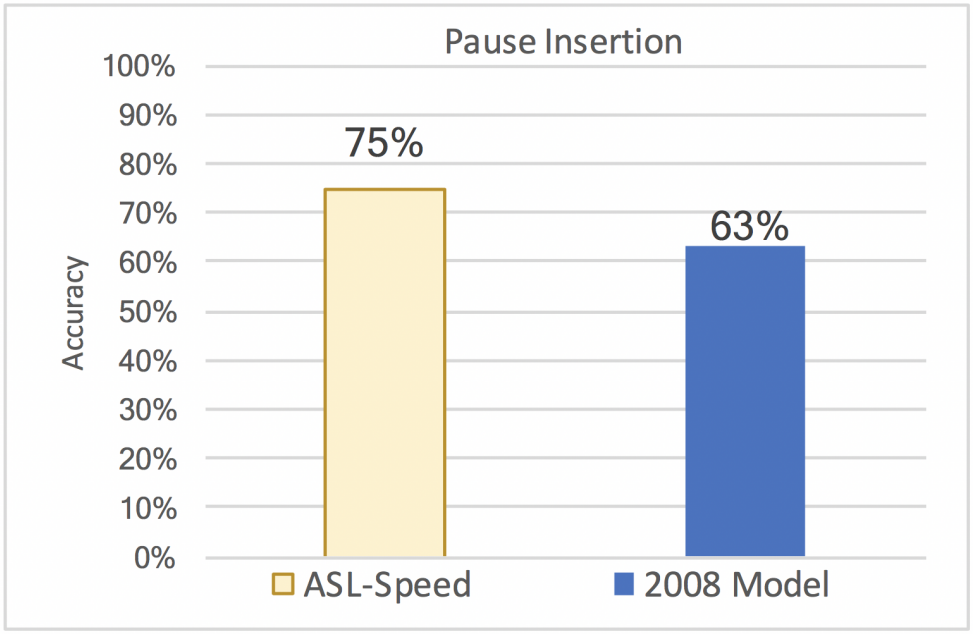


Figure 5.2: Comparison of our new ASL-Speed model and the 2008 Model on the Pause Insertion task - for a subset of passages from [84] for which we added syntax annotation

Figure 5.3 shows the prediction accuracy for the two regression tasks: Differential Rate and Pause Duration - Lower values for Root Mean Squared Error (RMSE) indicates better performance. The new ASL-Speed models outperformed the 2008 Model. Once again, this is an important result, since the 2008 Model had previously been the state-of-the-art method for predicting these speed and timing values for ASL animation.

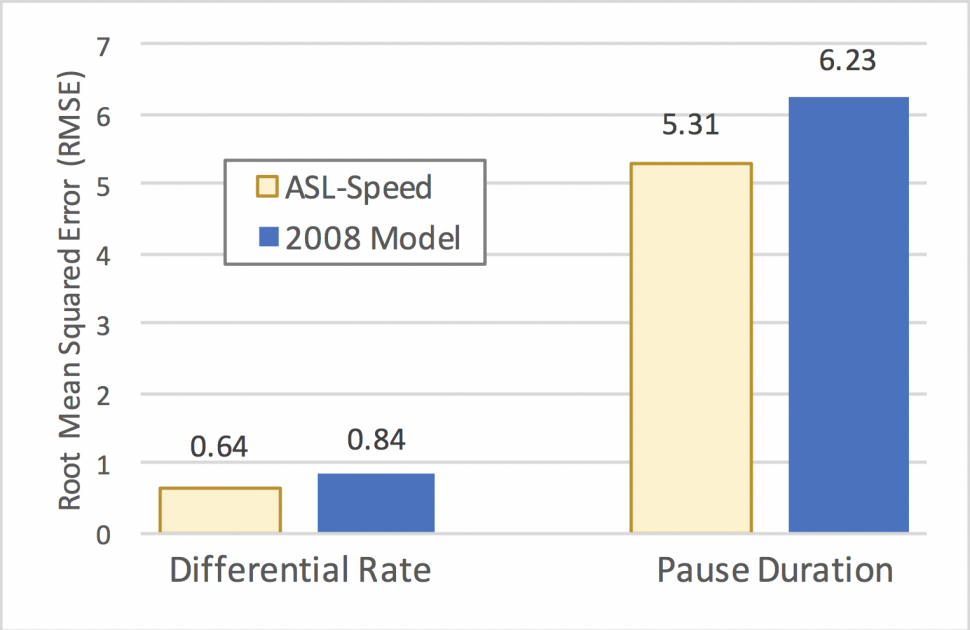


Figure 5.3: Comparison among new ASL-Speed model and the 2008 Model - for Differential Rate and Pause Duration

5.8 Conclusion

In this chapter, we presented our approach to engineer linguistic features for modeling some aspects of ASL animation generation. The ultimate goal was to find the best models that require a minimal set of features, in order to require as little input information as possible. As shown in Section 5.2, we divided our modeling tasks chronologically: pause insertion, differential signing rate, and pause duration. We presented our initial spot-checking comparison of various potential modeling methods, using a superset of our final set of features, to help us determine the most promising approach for each of the three tasks. Then, we performed parameter tuning to find the models with the best accuracy: Linear-Chain CRF is best modeling technique for pause inserting, and gradient boosting regressor is best modeling technique for differential rate and pause duration modeling. Our initial dataset-based evaluation of these three models was focused on a comparison of each model to a simple baseline. The baseline for pause insertion was a policy to insert a pause at the end of each sentence. The baseline for differential rate was a policy for each word to simply be performed at its average rate,

normalized to the fundamental signing rate of the signer. Finally, the baseline for pause duration was a policy to use uniformly assign a duration equal to the average pause duration in the corpus. Furthermore, we evaluated the quality of our models using dataset-based evaluation by comparing our robust models with the state-of-the-art rule-based 2008 Model of Huenerfauth [60, 63]. We found that our modeling approach had better results than the rule-based 2008 modeling approach. However, a simple dataset-based evaluation of the quality of ASL animation systems is not enough. Thus, there is a need for human-based evaluation of animations with parameters predicted by these models, as will be discussed in Chapter 6.

Chapter 6

Model Evaluation[14]

This Chapter presents our methodology and results during the user-based evaluation of our work. Specifically, we conducted a study with DHH participants (who are ASL signers with native fluency) to obtain first-hand feedback and judgements from these users about the quality and naturalness of the animations resulting from our model - along with more general input from these users about what aspects of speed/timing they deem most important when viewing animations of ASL. Overall, in this dissertation research, we are evaluating the quality of our modeling through three ways:

1. **Cross-Validation Model Evaluation**: In this method of evaluation, we are comparing our selected model with a proposed baseline. The goal of this evaluation is to select the robust model that will be used to build the animation. This method was discussed in detail in Chapter 5.

2. **Dataset-based evaluation comparing state-of-the-art modeling**: in the dataset-based evaluation we use our modeling to predict the timing parameter and compare that with the state-of-the-art rule-based approach. This is also discussed in detail in Chapter 5.

[14]The information in this chapter is based on a joint project with researchers in the Center for Accessibility and Inclusion Research (CAIR) at RIT (Larwan Berke, Sushant Kafle, Peter Yeung) supervised by my advisor (Dr. Matt Huenerfauth). The other researchers at CAIR assisted me with creating the study logistics and ASL videos recording. The user interview study of the project, which is the focus of this chapter, was presented at our paper at the ASSETS'18 conference [7], which received the best paper award and for which I was first author.

3. **User-based evaluation**: Since the goal of my research is to produce better animations of ASL for DHH people who use ASL, it is essential to conduct experimental studies of animation quality with DHH participants to learn their preferences about the animations generated with these models. This form of evaluation is the primary focus of this chapter.

In this chapter we address the fourth contribution.

Contribution 4: Empirically determine whether DHH ASL signers prefer animations of multi-sentence ASL passages in which timing values are determined by these new models or by the previous state-of-the-art rule-based technique.

6.1 User-Study with Subjective Feedback

While the metric-based evaluation in Chapter 5 is useful, it is important to test animations produced from these models with participants who are native ASL signers, to understand their reaction to these animations - and to ask for additional recommendations or feedback about how to improve them, as in [74]. For this evaluation, we chose an interview-based study design, in which a native ASL signer who was a researcher on our team met with native ASL signers to discuss their views on speed and timing in ASL animations, while looking at animations from our new model.

6.2 Participants

A total of 8 DHH participants were recruited on the Rochester Institute of Technology campus. Participants included 2 men and 6 women, of ages 21 to 34 (median age 23). Participants were native ASL signers: All participants learned ASL at their childhood (before age 3), 5 having Deaf parents, all of them having used ASL at school as a young child, and 5 with other Deaf family members. The researcher (also a Deaf native ASL signer) met the participants to conduct the interview in ASL.

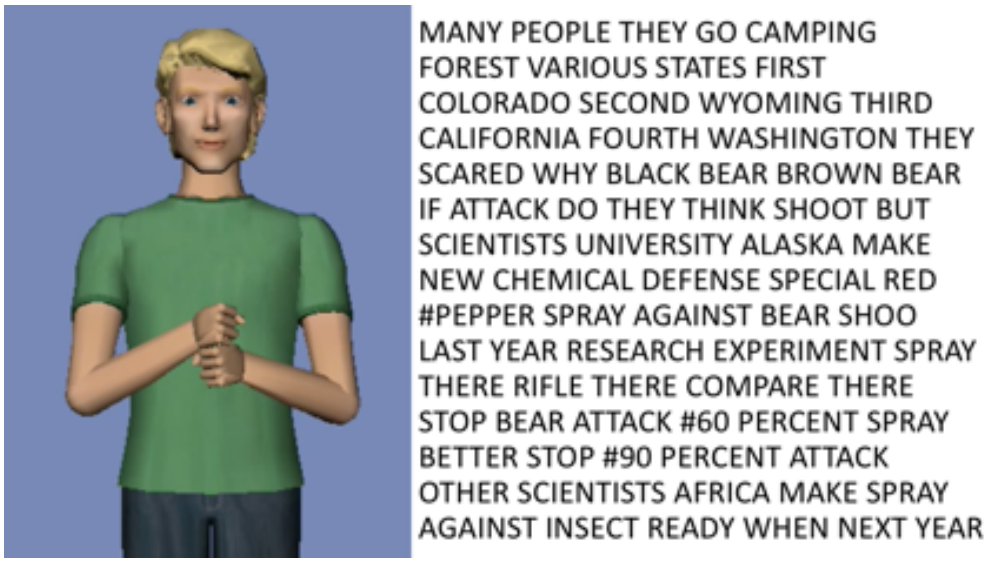


Figure 6.1: Image of animation (left) seen participants in the user study, and transcript (right). Participants did not see the transcript

6.3 Procedure and Data Collection

After some demographic questions, participants were asked to look at a laptop that displayed three pairs of animations of ASL (Figure 6.1). Each pair (shown side-by-side) was an identical passage, but one animation used speed and timing based on our new ASL-Speed model, while the other was based on the baseline models described previously, i.e. pauses at end of sentences and uniform sign speed and pause duration. The passages were selected from the stimuli shown in [60, 63], and each passage was approximately 75 words in length, on various topics (scientists developing a bear repellent spray, increases in rice prices, and a student selecting a career). Animations were generated using Sign Smith Studio [133], which allows for control of time parameters during animations, as described in [60, 63].

The semi-structured interview focused on participants' impressions of the animations' quality, which animation they preferred in each pair, and whether they noticed a difference between the two. When displaying the first pair of animations, we did not inform participants that speed or pauses differed; we wanted to see if participants noticed this on their own.

6.4 Feedback of the User Study

After the participants watched the stimulus animation, we asked the participants to answer some questions related to quality of the ASL animations (i.e. before showing the next one). These questions move from general questions to more specific questions. We divided our questions four areas: General questions, pausing questions, speed questions, and animation recommendation questions. For instance, some questions we asked the participants included:

- **General questions**:
  - What do you like about computer animations, in general?
  - What do you think about ASL animations?
  - Are there times when they would want more robotic signing or more natural signing?
- **Pausing-related questions**:
  - Do you think pauses and speed are important?
  - Is it easy to tell where the sentences begin and end?
  - How natural are the pausing of the animations? Why? Suggestion for improvement?
- **Speed-related questions**:
  - Is it too fast? Too slow?
  - How natural are the speed of the animations in the video? Why? Suggestion for improvement?
- **Selecting best animation questions**:
  - Which one of the two videos do you like? Why?
  - What difference did you notice between the two videos you just saw? How exactly?

When we asked a general question about computer animations, the participants found the idea of using animations to support ASL to be interesting, and they had a very high standard for what they would expect in regard to the quality of these animations. The participants indicated that they would expect the animations to look like Disney movies e.g. participant *P1* mentioned “Moana” from

the recent Disney movie as how they would expect the animation to appear. Furthermore, many participants indicated that they would expect the animations to appear realistic like a real person. While participants expressed some variation in their individual preferences for the body-type and colors to be used in a character performing sign language, there were some consistent answers to the following question: "What do you think is the most important characteristics for computer animation of ASL? Why?" All of the participants said that facial expression is the most important factor. Notably, prior to seeing animations during the study, none of the participants had initially mentioned the issue of speed or timing as being a concern about sign language animations.

After they had an opportunity to see the various side-by-side versions of the ASL animations (with one baseline and one ASL-Speed version in each case, in random left-to-right presentation), a majority of the participants (6 of 8) preferred the ASL-Speed animations, compared to the baseline one. Throughout the course of the interview, participants had an opportunity to view all three pairs of ASL animations, while commenting on these topics. The sequence of displaying the animations was rotated across participants, as well as whether the ASL-Speed animation was on the left or right.

Some comments were about the **overall speed, in general**:

- "*The [baseline] video was slow. If the animation signs slow, it means itself is a beginner or rookie. The [new] was fast. It means it's an experienced signer.*" (P2)[15]
- "*The [baseline] is good and has a good pace. The [new] is clearer than the [baseline]. It has strong ASL content than the [baseline]. It is clear, but fast.*" (P3)
- "*The [new] has a good pace. [Baseline] is fast.*" (P6)
- "*The [new] has a natural signing; almost similar to the real person signing.*" (P5)

Then, we asked them some questions that asked them to comment more specifically about the two animations' speed and timing, e.g. whether the boundaries between sentences were easy to perceive in the animations, whether the pauses in each animation seemed natural, whether the overall speed and timing of the animations seemed natural.

Some participants commented on the **pausing**:

[15] Participants did not know which animation was based on the baseline model or on the new model; they referred to the videos by pointing to them. We have edited their comments to indicate "*[baseline]*" or "*[new]*" for each participant (e.g. P2, P3, etc.). The comments were transcribed into English by the Deaf researcher.

- "*[Baseline] One minute of animation signing continuously - too overwhelming.*" (P1)
- "*[New] is fine with the pausing. It needs to improve quality. I suggest add a few pauses, but sign consistently.*" (P2)
- "*[New] is okay... Need to add 2 pauses. If no pause, I would be overwhelmed. That wouldn't be clear.*" (P3)
- "*I notice there are pauses [in the new]. The flow of the content is good and smooth.*" (P4)
- "*[Baseline] needs to add pauses between sentences, not continuing sign too fast. I lose information if it signs too fast.*" (P4)

One participant preferred the baseline animation in one pair, but he preferred the ASL-Speed in other pairs. In regard to the first pair, P6 commented "*It [baseline] has no pauses. It [new] pauses at the wrong time.*" For the pairs of animations for which the ASL-Speed was preferred, P6 said "*It [new] is normal, almost like a real person signing. It's better than the first one [baseline].*" and "*It's good [new]*"

Finally, some comments were about the **signing speed**:

- "*[Baseline] seems for beginners, and [new], for experienced signers... [New] is not too fast; it is a regular speed like a conversation... [New] is close enough to match the natural speed of an ASL signer.*" (P02)
- "*[New] is fine and steady.*" (P3)
- "*[New] is the right speed.*" (P5)

After viewing the animations, participants had other recommendations not specifically related to speed and timing. For instance, some participants addressed factors related to **animation colors and quality**:

- "*The person needs to wear black shirt, not a green shirt.*" (P1)
- "*Shirt need to be black. If the person is white, then he needs to wear black shirt. If the person is black, then he needs to wear a light gray shirt.*" (P1)

- "*Add the background, not too plain background.*" (P5)
- "*Color of the hand is hard to see. Does not match the shirt color.*" (P6)
- "*Background is not clear. Three bright (hands, shirt, and background). Need to balance color contrast.*" (P6)

Other participants have some feedback about the **avatar** performing the ASL:

- "*They need to put hands together on the middle of the chest.*" (P1)
- "*Increase size of the hand. Bigger the hands, easy to understand.*" (P5)
- "*Prefer different person, not white person every day.*" (P6)

Other participants shared some comments about facial expression:

- "*Good start, but the animations need to add more facial expressions.*" (P1)
- "*Facial Expression. Need to move the facial expression and the body. The body is too stiff and it looks like a robot. Again, it needs to move a lot of the body.*" (P5)
- "*Facial Expression is important because we can understand better. If there is no facial expression, it is hard to connect and hard to understand.*" (P5)
- "*Facial Expression. I feel the same as the real person. If it is bad, then the facial expression is too bland.*" (P6)

In fact, prior ASL-animation research at our laboratory had investigated the importance of facial expression, with several prior studies in this area [82, 87].

Returning to the issue of speed and timing, we noticed that individual participants may differ in the speed they prefer to receive animated signing, e.g. due to language-fluency, demographic, or experience factors, as discussed in [84]. In our interview study, 2 participants preferred a baseline animation - they perceived the baseline as having no pauses, and new as having incorrect pauses, and preferred lack of pauses to incorrect pauses. This insight suggests that we should consider the precision/recall tradeoff of our pause insertion model. The other 6 participants preferred ASL-Speed animations - commenting positively about speed/pausing.

While this was a relatively small interview study, the comments indicated a sensitivity to speed and pausing in these animations (as had been found in [60]), with a preference for the new model over the baseline. As discussed previously, this baseline (insert pauses at end of sentences only, use uniform signing rate and uniform duration for pauses) is how nearly all current sign-language animation systems generate their speed and pausing values.

6.5 Conclusion

This chapter provides additional evidence, from another form of evaluation, that our machine-learning models have out-performed the proposed baseline. Specifically, the results of this user-based evaluation study revealed that our participants preferred animations with the speed and timing parameters based on our new ASL-Speed models.

EPILOGUE FOR PART II

In part II we presented our methodology for building and evaluating new ASL animations based on a data-driven approach. In Chapter 5 we discussed the model-building process, with our methodology for selecting the best subset of features, how we tested different machine learning algorithms, and how we tuned the model parameters until we obtained a robust model. Model selection was based on comparing the best performing model with the baseline. We addressed how we evaluated our work using different types of evaluation. First, we compared our best performing model with the state-of-the-art rule-based approach. Moving to Chapter 6 we presented the second evaluation, by conducting interviews with DHH participants to understand their opinion about the generated ASL animations.

Part III: User Preferences for Speed, Timing, and Acceleration in ASL

PROLOGUE TO PART III

In this dissertation thus far, we found that a data-driven approach (trained on human recordings) can reveal patterns for timing parameters in ASL signing, enabling modeling of these phenomena, and these models can be used to synthesize natural ASL animations. However, there was an assumption built in to this prior work, namely that DHH users would prefer for an animation to move with speed and timing parameters *similar* to those of a human signer. In Part III, we have conducted a variety of user studies with human participants to investigate DHH users' preferences about speed and timing parameters, to evaluate this assumption.

Chapter 7 presents how we selected the appropriate timing parameters to be investigated. We conducted a pilot study with Deaf participants to identify the preferred range of values for five-speed and timing parameters including: sign duration, transition, differential signing rate, pause length, and pausing frequency. Then we conducted a user study to identified two preferred values from among five options for each parameter, one of which included a typical human value for this parameter. Finally, we conducted another study with Deaf participants to identify the most preferred value.

In Chapter 8 discusses the investigation of different speed profiles in the movements of ASL singing. We have examined prior work on acceleration curves of human movements in other signed languages, and then we have examined whether there are similar acceleration profiles in ASL. Finally, we have conducted a user-based study to evaluate whether participants prefer ASL animations in which the virtual human character moves in a manner that follows the typical acceleration curves seen in recordings of human signers.

Part III will address the following contributions:

Contribution 5: There is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation - for instance, prior work had found that users may prefer animations to be slower than human videos [63, 69]. Thus, for each of the timing parameters for ASL animation, we empirically determined which values of that parameter are preferred by Deaf ASL signers via an experimental study, in which animations with a range of such values are displayed for comparison.

Contribution 6: Prior research on speed and timing of sign-language animation has not specifically investigated the issue of predicting acceleration curves for the movements of the character's body [7, 63, 69]. Further, some prior linguistic research has observed different classes of acceleration curves used during or between words in French Sign Language [36], but such an investigation has not been performed for ASL. Thus, we examined our new dataset to conduct an analysis of motion-capture patterns of human movements, to empirically determine whether there are common categories of acceleration curves present in different linguistic environments, e.g. within ASL signs, between ASL signs, or near sentence boundaries.

Contribution 7: Following the same logic as for Contribution 5 above, since there is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation, we empirically determined whether accuracy in the use of particular acceleration curves influences the subjective judgements of Deaf ASL signers, as to the quality of the ASL animation.

Chapter 7

Empirical Investigation of Users' Preferred Timing Parameters for American Sign Language Animations

In the dissertation research described thus far, we have investigated the various ASL timing parameters using a data-driven approach. We found that the data-driven approach (trained on human recordings) can help to find interesting patterns for timing parameters and these patterns can be used to build natural ASL animations. However, there was an assumption built in to that prior work: specifically, that DHH users would prefer for an animation to move with speed and timing parameters similar to those of a human. Instead, perhaps DHH signers would prefer to receive the animations in a different timing configuration (for example, with more exaggerated or slower timing). In fact, there is a basis for this concern: Some prior research has found that DHH participants prefer a slower speed of animations compared to videos. So, we need to investigate whether DHH viewers actually prefer there to be a difference in the timing parameters for animation and for those

from human recordings, for different components of this speed, including for rate of pausing, the duration of pauses, the transitional movement time between words, and other timing parameters.

In this chapter we are presenting our empirical investigation of users preferred timing parameters, which will include the following specific goals:

- We empirically investigate the user preferences for each of five-timing parameter of ASL animation.
- After determining the general range of values that users prefer for each of these parameters, we conducted follow-up study to compare participants' preferences among the top two parameter values from the initial study. Notably, in the case of each parameter, a value based on human recordings was among the top two.

The primary goals of this chapter are to investigate the fifth contribution of this dissertation research:

Contribution 5: There is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation - for instance, prior work had found that users may prefer animations to be slower than human videos [63, 69]. Thus, for each of the timing parameters for ASL animation, we will empirically determine which values of that parameter are preferred by Deaf ASL signers via an experimental study, in which animations with a range of such values are displayed for comparison.

Chapter 7 will be divided to the following main sections, Section 7.1 will summarize some of the previous work relevant to this chapter. Section 7.2 will present some necessary definitions for this work. Section 7.3 will discuss a pilot user study conducted to confirm the methodology for investigating users' preferences for values among the five timing parameters. Next, Section 7.4 will present our Initial Five-Way Comparison Study, in which DHH signers compare animations based on timing values from human recordings to animations based on timing values identified in the pilot study in Section 7.3. After having identified the top two values for each parameter, Section 7.5 will present our Final Two-Way Comparison Study, to determine for each parameter whether users prefer animations based on a typical human value, or based on a value which is exaggerated in some way.

7.1 Prior Work Relevant to This Chapter

In this section, I will summarize some of the previous work discussed in Chapter 2 and Chapter 3, discuss additional related work relevant for this topic, and explain how that work inspired my current timing parameter investigation. Prior psycholinguistic research has investigated the timing and pausing parameters of both ASL and spoken English. For instance, Grosjean et al. conducted several studies on speed and timing of ASL based on observing video recordings of human ASL signers [43, 53, 55]. The five-parameter model of ASL speed and timing presented in Chapter 2 reflects the model of ASL that formed the basis for these researchers' work. One of their studies established that some timing parameters, e.g. the location and duration of pauses, are related to sentences' syntactic structure [53]. Overall, they found that human signers pause at approximately 25% of the inter-word locations during an ASL passage. They also found that longer pauses take place at sentence boundaries (approximately 229 milliseconds), with shorter pauses at other phrase boundaries.

Computational linguistic researchers have studied, using rule-based approaches, how to control an animated character to produce ASL with natural pauses and timing behavior. Prior researchers built and evaluated rule-based methods of predicting various ASL timing parameters [60, 63]. Some prior experimental research on ASL animations [60, 63] investigated animations with speed ranging from 0.9 signs-per-second to 3.0 signs-per-second. While not specific questions those authors had set out to investigate, their findings suggested that users preferred ASL animations to be slower than videos of human signers. However, the study had not investigated users' preferences for each of the five timing parameters; it had only investigated a single overall signs-per-second speed parameter. No prior research study has collected subjective preference judgements about ASL animations of various timing parameter values.

Prior research has examined how to set these various parameters of sign-language animation, but an **assumption implicit in much prior research** is that users would prefer for the output to appear as similar to humans as possible. For instance, research has examined how to improve the visual appearance of elements of the face [70], add natural body movements to signing [103], or point to locations in space in a manner similar to humans [49]. Prior work has demonstrated how the speed and timing details of an animation are critical for making that animation understandable to ASL signers [61, 64]. Our published work has also established that speed and timing of ASL animation is complex: consisting of multiple parameters, e.g. pause frequency, sign duration, etc. [10].

Many prior sign-language animation systems have employed **visual exaggeration** in the appearance of cartoon-like sign-language avatars, e.g. with enlarged faces [1, 47, 119] or over-sized hands [5, 93]. As the face and hands convey important linguistic information in sign language, researchers may utilize these visual exaggerations to enhance clarity for the viewer. While traditional hand-drawn cartoon animation techniques commonly employ visual exaggeration, especially in the face of characters [111, 125], artists commonly employ exaggeration of movement as well. For instance, Disney's basic principles of animation [81] recommend strategic use of exaggerated movement to produce dynamic and engaging animation. Despite this, prior sign-language animation work has largely focused on replicating human speed and timing, e.g. by building AI models of human speed and timing values so that those models can drive the movements of an animated avatar [8]. Despite one prior study that suggested that signers prefer animations of ASL to be slightly slower than videos of human signers [64], no prior study has rigorously investigated whether ASL signers actually prefer for the various timing parameters of ASL to be similar to those of human performance – or whether ASL signers would prefer some **exaggeration of the speed and timing** of these animations for clarity. A lack of prior empirical studies to evaluate these subtle aspects of ASL animation among DHH signers may have motivated designers and researchers to make the simplifying assumption that users would prefer animations to be as similar as possible to human recordings. We therefore conduct studies with DHH participants evaluating animations, which have revealed that **this assumption is incorrect**.

7.2 Background

Our research utilizes a five-part model of speed and timing for sign-language animation introduced by prior researchers [8, 10]. As illustrated in Figure 7.1, the five key parameters which investigated throughout this work include: sign duration, transition time, differential signing rate, pause frequency, and pause duration. The figure presents a series of timelines, each of which depict some variation in one of the timing parameters – with the middle timeline B in each case depicting a natural value for the parameter, which is typical of human signing. The A and C timelines depict two extreme values for each parameter, shown here to illustrate the concept. The five parameters include two relating to pausing (length and frequency) and three parameters relating to other aspects of speed and timing, as follows:

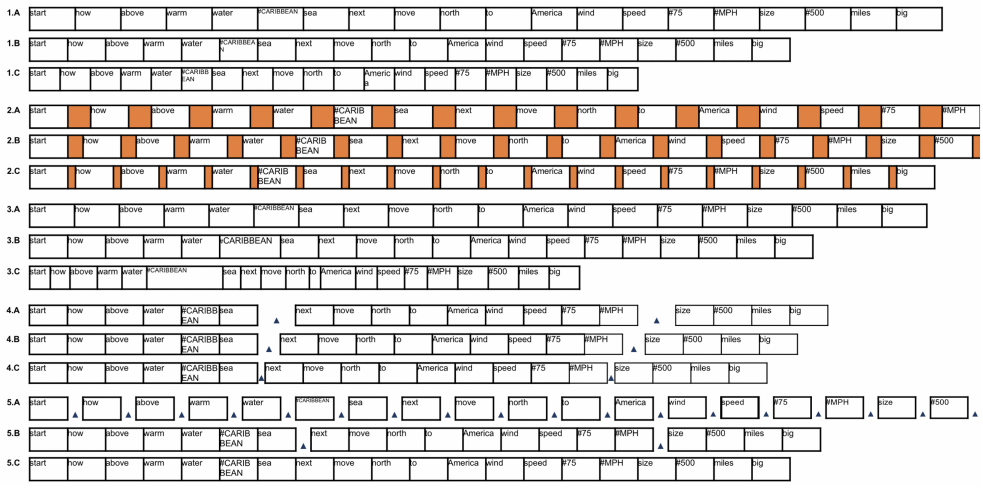


Figure 7.1: Visualization of five speed and timing parameters. The horizontal axis corresponds to a timeline representation of an ASL animation, with each rectangle representing the period of time of an individual word. (1) Variation in sign duration is illustrated in three alternatives A, B, and C, in which signs are produced at different speeds. (2) The transition time when the hands move from the final position of one sign and into the beginning position of the following sign may also be adjusted. (3) Signs may also be performed more quickly or more slowly due to various linguistic factors, which result in variation in differential signing rate, with more extreme speed-ups and slow-downs shown in the final timeline 3.C. (4) A signer may pause during signing for various linguistic reasons, and this timeline shows how the length of these pauses may vary. (5) The frequency with which someone may pause may also vary.

As shown in Chapter 3, prior work has found that adding linguistic pauses and adjusting the signing rate improves understandability of ASL animations, with DHH individuals sensitive to tiny errors in these parameters [60, 63]. In my recent work discussed in Chapter 4, Chapter 5, and Chapter 6, I have presented my approach of using a corpus of videos and motion-capture recordings of human ASL signers to build models to predict ASL timing parameters. In this chapter, since we are investigating users' perceptual preferences among animations, rather than building a model to incorporate withing an animation-generation pipeline, we discuss a slightly different set of timing parameters that can be directly observed in a final animation or recording of a human video:

- **Sign duration** is based on the original speed at which words were encoded in the animation system's lexicon and the overall speed at which the ASL animation is synthesized.
- **Transition** is time that a signer's hands are moving from the final position of one sign until the beginning position of the next, i.e. the time in-between two individual ASL words.
- **Differential signing rate** refers to how *dynamic* a signing performance is, i.e. the degree to which signs speed-up or slow-down due to various linguistic factors. For instance, words near the end of sentences may be performed more slowly, or the second appearance of a word in a conversation may be more quick [54, 56]. Differential signing rate is how extreme these variations in speed are during an ASL performance, and it may be conceptualized as an exponent applied to the speed-up or slow-down factor applied to each sign [10].
- **Pause length** is the time that a signer's hands stop moving between two ASL signs when performing a more substantial pause, which may be motivated by various prosodic or syntactic factors [54, 56].
- **Pausing frequency** is how often the signer pauses in ASL sentences, and it may be represented as the percentage of the inter-sign locations at which a pause occurs [10]. Of course, the distribution of such pauses is not entirely uniform, with more occurring at important syntactic boundaries, e.g. between sentences or phrases [54, 56].

The set of timing parameters in this chapter differs slightly from that in previous chapters in that we are now considering the transitional time in-between words, which had not been modeled in our earlier work. In addition, we are now considering these parameters in terms of observable values (e.g. average speed of hand movement, average time between words, percentage of inter-sign

boundaries with pauses). In our earlier work, since we were interested in creating models to insert pauses or adjust timing values as part of an animation synthesis pipeline, these same concepts had been expressed in an operational manner, i.e. manipulate the rate of a specific word, insert a pause at a specific location, etc.

7.3 Pilot Study[16]

While our research in previous chapters had investigated speed and timing models based on human recordings [7], earlier research suggested that users prefer ASL animations to be slower than human videos of ASL [60, 63]. Thus, in this pilot study we collect empirical judgments from native ASL signers for each of these parameters for ASL animation. In addition to providing us with guidance about whether the methodologies in this pilot study would be suitable in our upcoming studies, we also report specific values preferred by users in the pilot study, which has guided our selection of specific numerical speed and timing values investigated in the subsequent studies Section 7.3 and section 7.4.

7.3.1 Method

When we began to design our preliminary user study, we had to select a range of values for each of the timing parameters, so that we could show participants animations with various levels of each. In some cases, to select the "midpoint" of our scales for each parameter, we considered the typical speed and timing parameters used in other ASL animation systems. For instance, Sign Smith Studio (SSS) included some "default" values for various speed and timing parameters, and it provided the user with the ability to manually customize many of these timing parameters. For example, SSS used 0.25 seconds as the default transition time between words and 0.5 seconds as the pause duration at a sentence boundary.

For our IRB-approved study, a native-ASL-signer researcher on our team met in-person with native ASL signers on the Rochester Institute of Technology (RIT) campus to obtain their subjec-

[16] The information in this section is based on a joint project with my advisor (Dr. Matt Huenerfauth), I collaborated with a graduate student at RIT (Becca Dingman) whom assisted me with creating the ASL study logistics. The information on Section 7.3 was presented at our paper at the CHI'20 conference [9]

tive preferences on ASL animations which varied according to each of the five timing parameters. Appendix B illustrates the study stimuli used for this user study.

7.3.1.1 Participants

At the beginning of the appointment, which was conducted in ASL, the participants answered demographic questions. Our 16 participants included 8 women and 8 men, with median age 22.5 years old (range 18-25). Thirteen identified as Deaf, 1 as hard of hearing, and two were hearing children of Deaf adults who grew up using ASL since infancy. All participants learned ASL in childhood (age 2-8 years). Five had Deaf parents, 9 had Deaf family members or relatives. Fourteen of the participants used ASL at school as a young child. Three currently used only ASL at home, and 13 used both ASL and English at home. Fifteen used ASL at their college or university.

7.3.1.2 Procedures

On a 15-inch laptop, participants viewed a stimuli page. Figure 7.2 shows the general organization of the animations used in the page, while Figure 7.3 illustrates a zoomed image for one of the animations used in Figure 7.2. Our stimuli consisted of time-parameter-adjusted versions of a set of ASL passages which we had used in prior work [7] (see Appendix B.1 for the actual text of the stories), with each passage approximately 75 words in length. Sign Smith Studio (SSS) [133] was used to generated ASL animations [60, 63], and a researcher adjusted the timing parameters in the animations to produce stimuli with particular timing properties. Each screen displayed five variations side-by-side of an ASL animation of the same message, with each version using a different value for a particular timing parameter. For instance, one screen displayed five ASL animations with different values for the “sign duration” parameter. A screen like this appeared a total of five times, with each screen showing animations that varied based on a five timing parameters, e.g. transition time, pause duration, etc. Our study design, which included animations only as stimuli, was informed by prior methodological research published at ASSETS [102] and TACCESS [86] which had studied the impact of including a video of a human signer as a "topline" in a study; that prior work had recommended studies with animation stimuli only for comparison when differences are subtle.

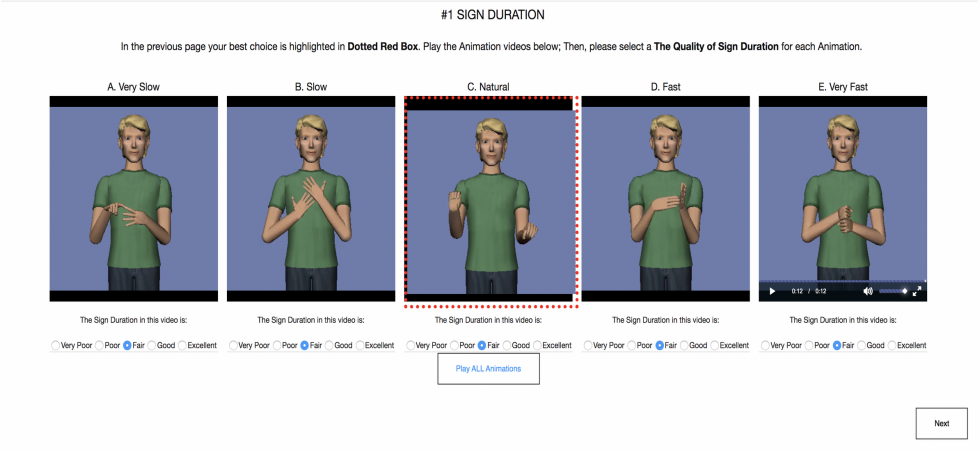


Figure 7.2: The interface for this study, which displayed five ASL animations of the same passage side-by-side, with each based on a different level of a particular timing parameter. Users indicated a scalar preference score (1 to 5) for each animation

Figure 7.3: A detail image of a zoomed-in region of Figure 7.2, showing one of the five ASL animation stimuli on the screen

Because the appearance variations of each animation can be somewhat subtle, some initial pilot testing with participants prior to launching the full pilot study revealed that users found it less confusing if the animations shown side-by-side were not displayed in randomized order: instead, those initial participants found it easier to compare the animations when they were presented in an arrangement with slower animations on the left and faster on the right. Initial testing also revealed that participants preferred to know "what was different" among the five variations displayed on a screen. Thus, prior to each screen of the study, users were informed of how the animations would vary, e.g. "On the next screen, you will see 5 animations in which the amount of time in-between words is different. Please evaluate each animation to indicate how you rate its quality."

For each of the timing parameters (sign duration, transition, pausing frequency, pausing length, and differential signing rate), after the participant provided their quantitative scores for the five side-by-side animations, the process was repeated, in a counterbalanced manner, for the next timing parameter. Another stimuli page was shown with five variations of an animation to collect judgements for the next parameter. Thus, each participant saw 25 videos during the study, providing numerical scores for each on a 1-to-5 scale (1 very poor to 5 excellent).

7.3.1.3 Results

Figures 7.4 to 7.7 and 7.9 show the average subjective score rating of participants for animations for each parameter. The Y axis in these figures represents the (1-5) user rating, while the X axis represents the five values of the timing parameter that had been displayed side-by-side.

We did perform statistical difference testing on users' subjective scores. Specifically, we analyzed responses using a Kruskal-Wallis test for each of the parameters, and we found a significant main effect for each parameter. We then performed post-hoc pairwise comparisons using a Wilcoxon test (with Bonferroni correction for multiple comparisons), to compare responses for each pair of levels for each parameter. Results of these tests are indicated as follows: In Figure 7.5, the "**" indicates statistical significance two levels ($p<0.01$); no other pairs in that figure had significant pairwise differences. In Figures 7.4, 7.6, 7.7 and 7.9, since nearly all pairs were significantly different, it was simpler to mark only those pairs which did were not significant ("n.s.").

For each parameter, the values shown in the study were selected as follows: First a "neutral/default" value was selected for each value (which was used as the middle option in each graph). Next, two slower and two faster variations were selected, so that five levels were investigated for each timing parameter. More details about each parameter and its values are below.

We compared five levels of sign duration (Figure 7.4): The middle level was 1.62 seconds, which was the average duration of signs in the ASL animations generated using SSS [133], using default timing of signs from the system's dictionary. We then quartered, halved, doubled, and quadrupled this value, to produce animation stimuli with average sign durations ranging from 0.41 seconds to 6.48 seconds. Larger sign duration values yield slower animations. Participants preferred the default 1.62 seconds/sign average timing. Thus we identified the range of values for sign duration parameter.

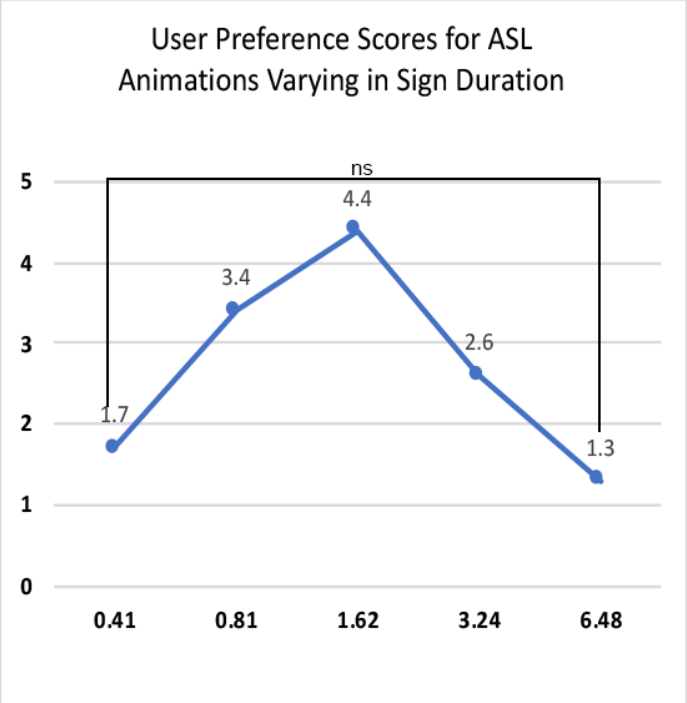


Figure 7.4: User preference scores for five animations which varied in their average Sign Duration (in seconds). All pairwise differences are significant except between the pair marked with "n.s."

We compared five levels of transition time (Figure 7.5): The middle value was 0.25 seconds (which was the default transition time between words in SSS), and the other stimuli used values of 0.125, 0.25, 1.0, and 2.0. Participants preferred the 0.25- or 0.5-second stimuli (no significant difference between these two). This result motivate us to focus on the narrower range of transition parameter that we need to work in in the upcoming experiments.

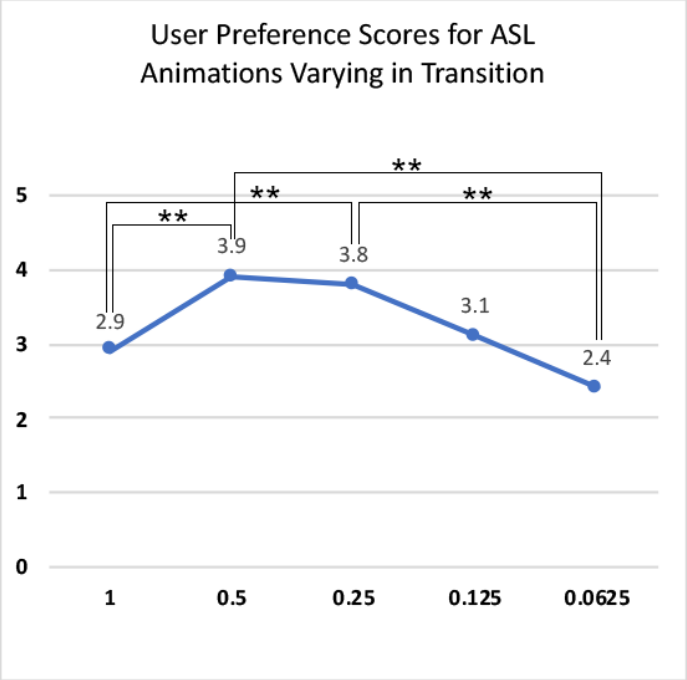


Figure 7.5: User preference scores for five animations which varied in their average Transition Time Between Signs (in seconds). Pairwise significant differences are marked with "**" ($p<0.01$)

We compared five levels of pause duration (Figure 7.6): The middle value was 0.5 seconds (the default time of a pause between sentences in SSS), and the other levels included: 0.125, 0.25, 1.0, and 2.0. Users preferred animations with pause duration of 0.5, 0.25, or 0.125 (no significant difference between these three levels). In the upcoming studies, we will focus on this non significant difference.

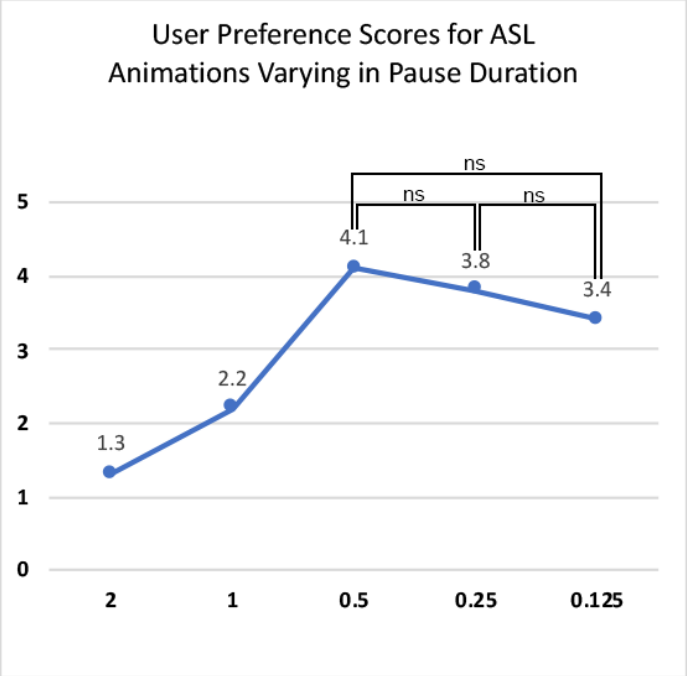


Figure 7.6: User preference scores for five animations which varied in their average Pause Duration (in seconds). All pairwise differences are significant except between pairs marked with "n.s."

The pausing frequency parameter (Figure 7.7) is how often the signer pauses (e.g. stops moving between phrases or sentences), represented numerically as a percentage of the between-sign locations where a pause occurs. However, inserting pauses randomly among these locations would have yielded an unnatural result; so, we used the following rubric to define our five levels for this parameter in the stimuli: 0% (no pauses inserted), 14% (pauses inserted after every sentence), 31% (same as previous, plus pauses inserted after every clause and verb phrase), 49% (same as previous, plus pauses after every noun phrase), and 100% (pauses inserted after every word). Users preferred animations with a pause after every sentence, which, in the stories shown in our study, was at 14% of the inter-sign locations.

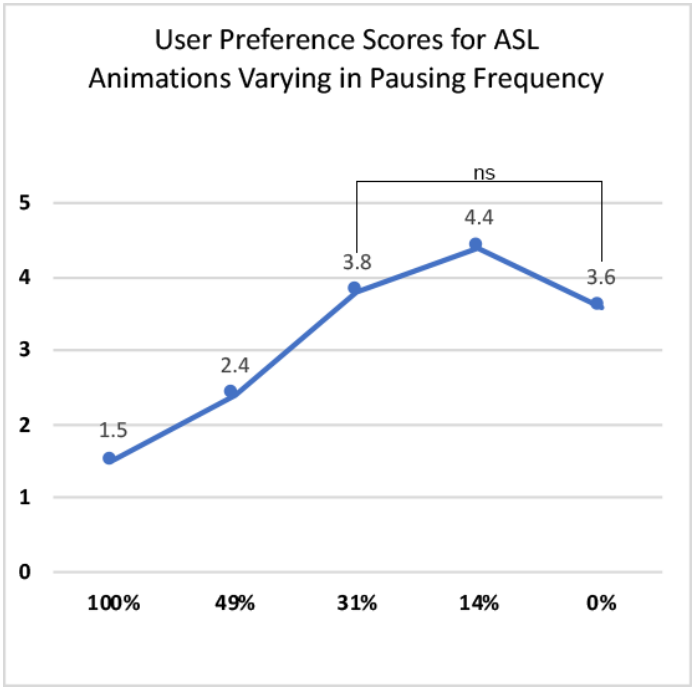


Figure 7.7: User preference scores for five animations which varied in their Pausing Frequency (represented as the percentage of inter-sign gaps where a pause occurs). All pairwise differences significant except one marked

The differential rate parameter (Figure 7.9) represents the tempo dynamics of an ASL signing passage, in which signers vary their speed throughout a passage, e.g. slowing down at the end of sentences. As we discussed in Section 5.5, we trained a Gradient Boosting Regressor model on motion-capture recordings of human ASL signing [100]. Our resulting model can predict, for each word in an ASL passage, a "speed adjustment" factor that should be used to adjust the speed of an individual ASL sign, based on some surrounding linguistic properties around that word, e.g. how close it is to the end of a sentence. Thus, the value of the differential rate parameter in this study represents the power ("Exponent" in Equation 1) to which we raise the speed adjustment factor for this word (as predicted by our model). Thus, in our prior work when we developed this model, we used a value of 1 for this Exponent; essentially, we had directly used the output of the model as has been trained on the tempo dynamics of human ASL signers in our corpus. By adjusting this Exponent, we can dampen or magnify the effect of this speed adjustment factor, to produce animations that are more consistent or more variable in their signing speed. In our stimuli, we presented users

$$V_{final} = V_{original}.Factor^{Exponent} \quad (7.1)$$

Figure 7.8: The final velocity (vfinal) of a sign is based on its original velocity (voriginal) multiplied by a speed adjustment factor (Factor), which may be raised to some power (Exponent). In Figure 7, the values shown along the x-axis represent this Exponent

with animations that used an Exponent value of 0.25, 0.75, 1, 1.5, or 2. We found that participants preferred animations that had differential rate similar to human signers, i.e. with an Exponent of 1 for the speed adjustment factor, which was based on our model trained on human recordings of ASL.

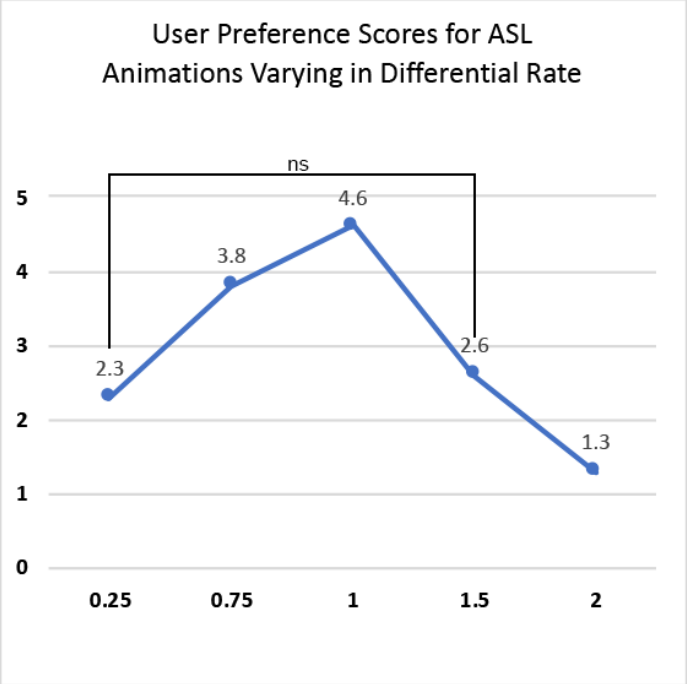


Figure 7.9: User preference scores for five animations which varied in the exponent used when applying their Differential Rate factor (see Equation 1). All pairwise differences significant except pair marked with "n.s.". The typical value for differential rate located at "1" which corresponds to the differential rate variability based on humans, with higher values reflecting more exaggerated variations in speed.

The goal of this pilot study was to identify the general range of values which might be the most preferred by participants so that specific values could be selected for investigation in subsequent studies.

7.4 Initial Five-Way Comparison Study[17]

[17] The information in this section is based on a joint project with my advisor (Dr. Matt Huenerfauth), I collaborated with post Ph.D. student Sooyeon Lee and graduate student at RIT (Becca Dingman) whom assisted me with creating the ASL study logistics and paper writing. The information on Section 7.4 was presented at our paper at the ASSETS'21 conference [**?**]

From the outcome of the preliminary **Pilot Study** (Section 7.3) we know the general range of values which users preferred, in the **Main Five-Way Comparison Study** we selected levels for timing parameters that "zoom in" on the most-preferred range.

7.4.0.1 Difference Between User Studies

Our pilot study had demonstrated that our overall methodology was effective for presenting stimuli and gathering users' preferences for each of these individual timing parameters of ASL animation. Our pilot study had systematically investigated users' preferences for five timing parameters individually. Specifically, in a comparison of five levels for each parameter, our pilot study had identified some preferred values for sign duration, pausing frequency, and differential rate. In the case of transition time and pause duration, our pilot study had only revealed significant differences between some pairs of values, but it had not identified a single preferred value for those two parameters, among our five stimuli. The initial findings of our pilot study have informed our selection of a set of values which are the focus in our main Initial Five-Way Comparison Study.

- In order to examine users' preferences for each timing parameter individually, to produce a set of five animations that depicted variation in the value of that one parameter, in the Pilot Study we had to select an **arbitrary value** for the remaining four parameters, which were held constant across the five animations displayed side-by-side [10]. Without prior empirical basis for selecting the value of those other four parameters, we selected sub-optimal values, a possible confound. In Initial Five-Way Comparison Study and Final Two-Way Comparison Study, when generating a set of animations to investigate the one parameter of focus, we use the preliminary findings from the Pilot Study to select "default" values for other parameters.

- Similarly, when selecting the range of values to investigate for each parameter, we did not have a prior empirical basis. In some cases, we selected a range of values to compare which was too broad, which raises the possibility that users may have actually preferred a value that was in-between two values they evaluated. In our Initial Five-Way Comparison Study, we **narrow and re-center the range** to obtain better precision.

- When selecting the set of values to compare for each parameter, our Pilot Study had not included a value based on the **typical value in human ASL signing**, instead our goal was to determine the users' preferred range for each timing parameter. However, we still need to

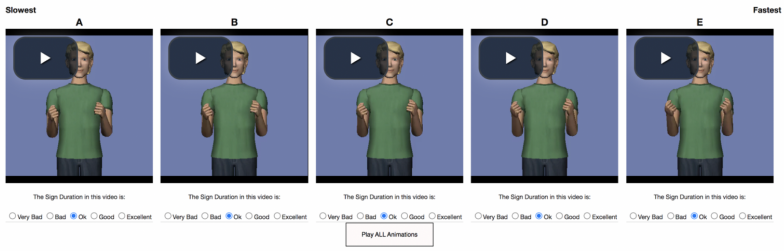


Figure 7.10: The interface for *Final Two-Way Comparison Study*, which displayed five ASL animations of the same passage side-by-side, with each based on a different level of a particular timing parameter. Participants subjectively evaluated each animation on a 5-point scale.

determine whether participants' subjective rating of an animation with their most preferred timing value differed significantly from their rating of an animation with a parameter value based on a human. Our Final Two-Way Comparison Study specifically compares participants' most preferred animations, including one with human-based timing, to settle this issue.

7.4.1 Method

In this study, we presented participants with multiple animations to identify which values of each timing parameter resulted in animations that obtained the highest subjective ratings from DHH ASL signers. Our methodology is built upon the previous Pilot Study, as described in sub-subsection 7.4.0.1. When designing the current study, we addressed several limitations of that pilot study, by: selecting more optimal default values for parameters, re-centering and narrowing the range of values evaluated, and directly comparing the most preferred animation to an animation based directly on human timing values. The goal of the current Initial Five-Way Comparison Study is to identify the top-two preferred values for each speed and timing parameter in ASL. As will be discussed in the results below, for each parameter, the top two always included a value that is typical of human signing.

In this IRB-approved study, which was conducted via online videoconferencing due to the COVID-19 pandemic, a researcher on our team who was a native ASL signer met with participants to obtain their subjective preferences about ASL animations, which varied according to each of the five timing parameters. During the appointment, the participant was provided a link to a website that had multiple pages, one for each timing parameter: sign duration, transition time, differential signing rate, pause length, and pausing frequency. The website was designed so that the order of

these five pages was randomized. Each page contained five ASL animations side by side like Pilot Study, all displaying the same ASL message, but with each animation varying in the value of a specific timing parameter, e.g. sign duration, as shown in Figure 7.10. Other-than conducting this study in online video conferencing, for this study, we followed the same methodology of a prior preliminary Pilot Study.

7.4.1.1 Participants

At the beginning of the appointment, the participants answered demographic questions. Our 20 participants included 11 who self-reported a “female" and 9 “male." Their ages ranged from 19 to 37 years old, with a median age of 24. Sixteen participants identified as Deaf, and four, as hard of hearing—with all participants having this auditory status since infancy or early childhood. Seven participants had Deaf parents, 12 used ASL with their parents since infancy or early childhood, and 18 used ASL as a young child at school. All participants reported using ASL at college or university.

7.4.1.2 Stimuli

As described below and displayed in Figure 7.11, for each timing parameter, the middle value selected for each parameter was based on typical human signing, with values that were double or half used as the upper or lower bounds in most cases in the prior Pilot Study, with the ranges reduced in this current study based on those results. We used the findings from that prior preliminary Pilot Study to select the overall range of values for each timing parameter and to select reasonable “default" values for the other timing parameters, when creating stimuli for a particular parameter.

- **Sign Duration:** A prior analysis [10,101] of specific words (matching those in our stimuli) in a corpus of human signing recordings revealed that the typical human level for this parameter is for an overall passage to have an average sign duration of 1.28 seconds. We selected the other four levels (3.24, 1.62, 0.81, and 0.41 seconds) to be centered on this human level as explained above, to make it more likely that the range of stimuli values bracket the anticipated preference of users, while providing sufficient granularity. Smaller sign duration values yield faster animations. We first constructed stimuli with sign durations based on the standard base duration of each sign in Sign Smith Studio [134]; next, these durations were re-scaled to produce an animation with desired average sign duration, to produce the stimulus corresponding to each level.

- **Transition Time:** The typical human level for this parameter was 0.23 seconds, as calculated in prior work from a corpus of human signing [10, 101]. We selected the other levels for this parameter through similar considerations as above: 0.0625, 0.125, 0.5, and 1.0. Larger values indicate slower transitions between words. We produced each animation stimulus by adjusting the transition time in-between words to each of these levels.

- **Differential Signing Rate:** As described above, this parameter represents the degree to which speed-up or slow-down factors are applied, to adjust the speed of individual words in a passage, based on linguistic context. In our study, to obtain this speed-up or slow-down factor for each word, we used a prior model which had been trained on human recordings [101]. As this parameter is represented as an exponential weighting of the factor, the typical human level can be represented by an exponent of 1. Following considerations like those for other parameters above, we selected the other four levels as 0.25, 0.75, 1.5, or 2. Higher exponents will result in more extreme temporal dynamics, and exponents less than 1 result in less extreme variations in speed for each word. We produced stimuli by applying this exponential weighting to the speed factor predicted by the model [101].

- **Pause Duration:** The typical human level for the duration of pauses is 0.22 seconds, based on an analysis of a corpus of human recordings [10,54]. In this study, we began by predicting the duration of each pause, using a prior model trained on human recordings [101]; these pause lengths were then re-scaled to produce animations with desired average pause-length duration, to produce stimuli. We selected the other levels for this parameter through similar considerations as above: 0.0625, 0.125, 0.5, and 1.0. Larger values indicate longer pauses.

- **Pausing Frequency:** As described previously, this parameter represents the numerical percentage of the intervals in-between words where a signer performs a longer, prosodic pause. To select levels for stimuli, we did not wish to randomly insert pauses during the ASL passage, to avoid an unnatural result. We therefore established policies for pause-placement to obtain five levels: 0% (no pauses inserted), 14% (pauses inserted after every sentence), 25% (pauses inserted after every sentence and every clause), 49% (pauses inserted after every sentence, clause, verb phrase, plus pauses after every noun phrase), and 100% (pauses inserted after every word).

Parameter	Levels in Study 2	Very Bad	Bad	OK	Good	Excellent	% Very Bad % Bad % OK % Good % Excellent	Mean	Post Hoc Wilcoxon (p<0.05)	Study 3
Sign Duration Values for each level represent average time spent during signs, across entire stimulus, in seconds.	a. 3.24 (slow)	60%	10%	30%	0%	0%		1.7	Worse than b, c, d	
	b. 1.62	5%	20%	45%	15%	15%		3.2	Better than a, e	
	c. 1.28 (typical human)	5%	5%	25%	55%	10%		3.6	Better than a, e	Yes
	d. 0.81	0%	15%	30%	35%	20%		3.6	Better than a, e	Yes
	e. 0.41 (fast)	35%	45%	15%	5%	0%		1.9	Worse than b, c, d	
Transition Values represent average time spent during between-sign transitions, across entire stimulus, in seconds.	a. 1 (slow)	15%	20%	30%	25%	10%		3.0	Better than e	
	b. 0.5	5%	5%	35%	45%	10%		3.5	Better than d, e	Yes
	c. 0.23 (typical human)	0%	10%	45%	40%	5%		3.4	Better than d, e	Yes
	d. 0.125	20%	30%	30%	15%	5%		2.6	Worse than b, c	
	e. 0.0625 (fast)	45%	15%	25%	10%	5%		2.2	Worse than a, b, c	
Differential Rate Values represent dynamic speed variation during signs, as exponent weighting of the speed-up factor.	a. 0.25 (less dynamic)	15%	35%	25%	10%	5%		2.3	Worse than b, c; better than e	
	b. 0.75	10%	0%	25%	35%	30%		3.8	Better than a, d, e	Yes
	c. 1 (typical human)	5%	10%	20%	25%	30%		3.4	Better than a, d, e	Yes
	d. 1.5	25%	25%	35%	15%	0%		2.4	Worse than b, c; better than e	
	e. 2 (more dynamic)	70%	10%	20%	0%	0%		1.5	Worse than a, b, c, d	
Pause Length Values for each level represent the average time spent during pauses, across entire stimulus, in seconds.	a. 1 (long)	40%	20%	30%	5%	5%		2.2	Worse than b, c, d, e	
	b. 0.5	10%	25%	25%	30%	10%		3.1	Better than a	
	c. 0.22 (typical human)	5%	0%	50%	25%	20%		3.6	Better than a	Yes
	d. 0.125	0%	5%	50%	40%	5%		3.5	Better than a	Yes
	e. 0.0625 (short)	5%	30%	25%	35%	5%		3.1	Better than a	
Pausing Frequency Values for each level represent the percentage of between-sign locations where pauses occur.	a. 100% (always)	60%	15%	15%	10%	0%		1.8	Worse than b, c, d, e	
	b. 49%	20%	30%	30%	10%	10%		2.6	Better than a, worse than c, d, e	
	c. 25% (typical human)	0%	5%	55%	25%	15%		3.5	Better than a, b	Yes
	d. 14%	5%	15%	35%	40%	5%		3.3	Better than a, b	Yes
	e. 0% (never)	10%	10%	30%	35%	15%		3.4	Better than a, b	

Figure 7.11: The five values selected for comparison in the Initial Five-Way Comparison Study (study 2 in the graph) and used as the basis for ASL animation stimuli for each parameter: sign duration, transition, differential rate, pause length, and pausing frequency. Participants evaluated each stimulus on a five-point scale, and the percentage of participants selecting each option (Very Bad, Bad, OK, Good, and Excellent) is shown, along with a divergent stacked bar graph redundantly visualizing these percentages. A mean score is also shown, calculated using 1 for Very Bad, 5 for Excellent, etc. Results of pairwise post hoc Wilcoxon tests comparing each level are presented, based on Bonferroni-corrected p-values ($p<0.05$), to explain how the “top two" values were identified for each parameter, which were subsequently compared in the Final Two-Way Comparison Study (study 3 in the graph).

7.4.1.3 Results

Figure 7.11 lists the five timing parameters evaluated during our Initial Five-Way Comparison Study, as well as the various levels of each parameter, i.e. the specific timing values used in the ASL animations, with the "typical human" level for each labeled. For sign duration, transition, and pause length, the values are in seconds. For pausing frequency, the values are percentage of signs after which a pause immediately occurs, and for differential signing rate, the value indicates the exponential weighting of the speed-up or slow-down factor. Figure 7.11 presents the percentage of participants who selected each option: Very Bad, Bad, OK, Good, Excellent, with a divergent stacked bar graph visualization of those percentages. The term "OK" was used as the mid-point on the scale in consideration of participants who may have lower English literacy who may be less familiar with the vocabulary word "Neutral". A mean score on a 1-to-5 scale is also presented, calculated using 1 for Very Bad, 5 for Excellent, etc.

We conducted statistical significance testing on users' subjective ratings of the animations using a Kruskal Wallis test, and we found that there was a significant main effect for each parameter. We conducted post hoc Wilcoxon Ranked tests between all pairs of levels for each parameter, and Figure 7.11 summarizes these pairwise results, based on Bonferroni-corrected p-values. In general, our goal in our Initial Five-Way Comparison Study was to identify a "top two" most preferred levels for each parameter, which we would later compare in our Final Two-Way Comparison Study. For some parameters, our statistical analysis only revealed a top three, and in such cases, we preferred to select adjacent values such that the typical human level was included in the two values that would be compared in our subsequent Final Two-Way Comparison Study.

- **Sign Duration:** Participants preferred an average sign duration of 0.81, 1.28, or 1.62 seconds ($X^2 = 41.775$, df=4, N=100, p<0.05); with no significant pairwise differences among these top three levels. As 1.28 is the typical human value, and since 0.81 received more Good or Excellent ratings than did 1.62, we decided to select 0.81 and 1.28 for inclusion in the Final Two-Way Comparison Study.

- **Transition Time:** Participants preferred the 0.23 and 0.5 levels ($X^2 = 17.145$, df=4, N=100, p<0.05), but post hoc testing did not reveal a significant pairwise difference between these two levels. We will therefore compare these two in the Final Two-Way Comparison Study.

- **Differential Signing Rate:** Participants preferred the 0.75 and 1 levels ($X^2 = 38.064$, df=4, N=100, p<0.05), but post hoc testing did not reveal a significant difference between these two levels. We will therefore compare these two in the Final Two-Way Comparison Study.

- **Pause Length:** Participants preferred animations with levels of 0.22 and 0.125 seconds ($X^2 = 16.197$, df=4, N=100, p<0.05), but post hoc testing did not reveal a significant difference between these two levels. We will therefore compare these two in the Final Two-Way Comparison Study.

- **Pausing Frequency:** Participants preferred animations with levels 25%, 14%, and 0% ($X^2 = 24.603$, df=4, N=100, p<0.05), but post hoc testing did not reveal a significant difference between these three levels. As prior work had established that humans typically pause at 25% of inter-sign locations [10], we will therefore compare the 25% and 14% levels in the Final Two-Way Comparison Study.

7.4.1.4 Discussion

This study identified two preferred values from among five options for each parameter, one of which included a typical human value for this parameter. Notably, for each parameter, an animation based on the typical human value was among the top two preferred values.

7.5 Final Two-Way Comparison Study

The goal of the Initial Five-Way Comparison Study had been to identify the most preferred two values for each of the five timing parameters. However, that study was not sufficiently powerful to identify which of these two values was the most preferred by users; so, the focus of our Final Two-Way Comparison Study is to determine, in the context of a side-by-side comparison, whether there is a measurable difference in participants' subjective rating of the quality of animations based on these top two values. Notably, for each parameter, an animation based on the typical human value was among the top two preferred values in Initial Five-Way Comparison Study; thus, Final Two-Way Comparison Study has a more fundamental goal, namely to determine for each parameter whether users prefer animations based on a typical human value, or based on a value which is exaggerated in some way.

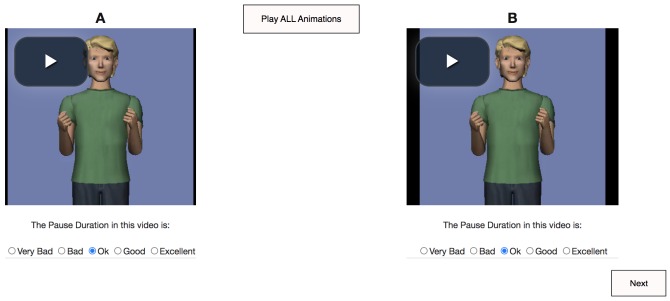


Figure 7.12: A sample screenshot for one of the pairs of animations displayed in the *Final Two-Way Comparison Study*.

The rationale for conducting this follow-up study with identical stimuli to the earlier study was that we were concerned that with five animations displayed side-by-side in Initial Five-Way Comparison Study – and with only a five-point scalar response item for participants to indicate their rating – we may not have had sufficient granularity available in the response variable. By presenting these two items side-by-side for comparison, we expected that participants may be more likely to use the range of the scalar response item to indicate a preference between the two side-by-side stimuli.

7.5.1 Method

The methodology in Final Two-Way Comparison Study was nearly identical to that of Initial Five-Way Comparison Study, with one exception: As shown in Figure 7.12, on each screen the participant was asked to compare only two animations side-by-side. As in Initial Five-Way Comparison Study, the website randomized the order of the screens so that the sequence of the five timing parameters was shuffled, and in Final Two-Way Comparison Study, the left-to-right arrangement of the two animations was also randomized.

7.5.1.1 Participants

The researcher conducting the study in ASL was once again a Deaf native ASL signer, who met virtually with the participants; this study also began by asking the 20 DHH participants to respond to demographic questions. Only 1 of these participants had also been a participant in Initial Five-Way

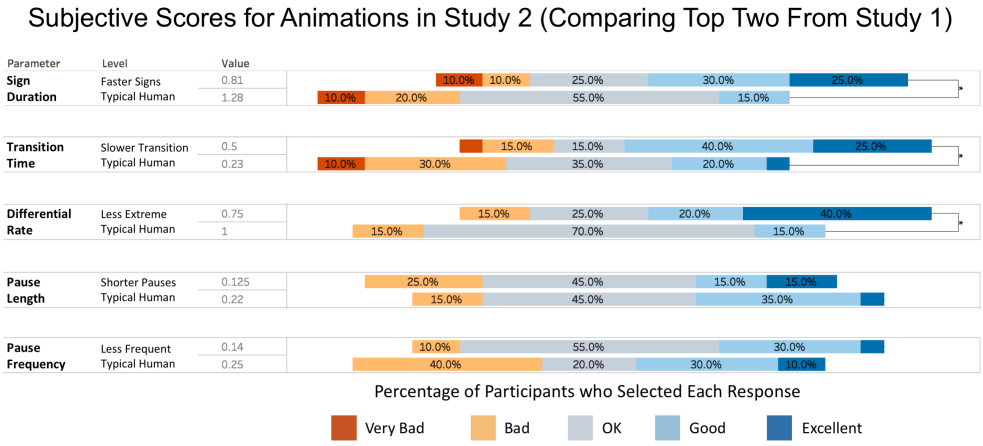


Figure 7.13: Participants' subjective ratings in *Final Two-Way Comparison Study*, which compared ASL animations with the top two levels from *Main Five-Way Comparison Study*, for each of the five timing parameters. The divergent stacked bar graph shows the percentage of participants who evaluated each animation as: Very Bad, Bad, OK, Good, or Excellent. For each parameter, significant pairwise differences are marked with * ($p<0.05$).

Comparison Study. Participants included 9 men and 11 women, of ages 19 to 56 (median age 27). Eighteen identified as Deaf, and two as hard of hearing. All but 1 participant had learned ASL before age of 5; the 1 remaining participant had learned ASL at age ten years old. Five participants had Deaf parents, and 12 participants had other Deaf family members, e.g. brothers or sisters. Seventeen participants used ASL in their home with their parents since infancy or early childhood.

7.5.1.2 Results

Figure 7.13 illustrates participants' subjective scores for the animations in the Final Two-Way Comparison Study. For each of the five timing parameters, to compare users' subjective scores for the pair of animations, a Wilcoxon Ranked test was performed. Significant pairwise differences ($p<0.05$) are marked with a star ("*") in figure 7.11. We draw a distinction between the 2 pausing-related values (for which participants preferred values similar to human signing) and the 3 non-pausing-related values (for which participants did not). The overall pattern of results can be best summarized by separately considering the speed-related and pausing-related timing parameters, as follows:

- For the sign duration, transition time, and differential signing rate parameters, a significant difference ($p<0.05$) was observed between participants' subjective scores for the two animations. For sign duration, participants preferred animations with average sign duration of 0.81 seconds, rather than animations with average sign duration of 1.28 seconds, which is the value in typical human signing. For transition time, participants preferred animations with average transition time of 0.5 seconds, rather than those with 0.22 sections, which is the value in typical human signing. For differential signing rate, participants preferred animations with a speed-up-or-slow-down factor raised to an exponent of 0.75, as compared to those raised to an exponent of 1, which would correspond to the speed dynamics of typical human signing. **Participants preferred animations with signs that moved slightly faster than human signers, with transitions in-between signs that moved slightly slower than human signers. They also preferred animations to exhibit less speed dynamics (with less extreme speed-ups or slow-downs during sentences) as compared to speed changes in human signing**.

- For the pause length and pausing frequency parameters, no significant difference was observed between participants' subjective scores for the two animations in each case. For each parameter, we conducted post hoc statistical equivalence testing, using the Two One-Sided Tests (TOST) method, to determine whether the scores for each pair of animations were indeed statistically equivalent. With a margin of 1 (on the 5-level scalar response item), the TOST revealed that participants had equivalent subjective ratings of animations based on either of the top two values observed in study 1 for pause length (p-value < 0.0017) and pausing frequency (p-value < 0.0008). **Overall, participants preferred ASL animations that used values similar to those of typical human signing both for both how often to pause and for how long the pause should be.**

7.5.1.3 Discussion

While in previous studies we presented a wider range of values (i.e. five levels) for each timing parameter, however, in the *Final Two-Way Comparison Study* we used a narrower range. Specifically, in this study, for each of the five timing parameters, we asked participants to simply compare only two animations side-by-side. In addition to identifying the single most preferred value for each

parameter, this study also enabled us to determine whether ASL signers prefer for animations to use timing values based on a human – or timing values that are exaggerated in some way.

7.6 Chapter Discussion

Prior approaches for selecting speed and timing parameters for ASL animations have included: asking human artists to set these details of the animation manually (considering video of a human performance), writing rules (based on linguistic research on human signing) to engineer speed and timing for ASL animations [61, 64], or building software to predict speed and timing based on AI models trained on human signing [8]. These approaches all assume that the speed and timing values for an ASL animation should be as similar as possible to the speed and timing values in typical human signing, and our findings reveal that, for certain ASL speed and timing parameters, **this assumption is incorrect**.

Individuals pursuing all three approaches above can **make use our findings** to improve DHH users' satisfaction with the resulting ASL animations. For computer-animation artists who seek to create realistic ASL animations, our findings suggest that rather than directly copying the timing of a video recording of a human, it may be more effective to adjust the sign durations, transitions, and differential rate to differ from typical human levels. For researchers investigating how to automate the synthesis of ASL animations, our findings may inform how the they approach the speed and timing of the animation affects in their work. For researchers using rule-based approaches to set these parameters, it may be valuable to adjust numerical values in their rules so that the output animation aligns with our recommended parameter values. Researchers training AI models to predict speed and timing may need to apply some correction factor to their model predictions or modify timing values within their dataset prior to training.

While **exaggeration of both appearance and movement** is common in hand-drawn and computer animation [81, 111, 125], prior work on sign-language animation has generally attempted to maximize the realism or similarity to human signers. In particular, prior work on determining the avatar movement has focused on replicating as closely as possible how human signers move, with prior work on the use of the signing space [72], realistic facial expressions [68], movement paths of the hands during complex verbs [71], and—most relevant to our work—speed and timing parameters [3, 8, 64]. The appearance of avatars has also reached new levels of realism in recent years, e.g. [135]. However, compared to movement, avatar *appearance* has seen more examples of

utilization of exaggeration, e.g. through enlargement of the the face or hands [1, 5, 47, 93, 119], either for greater clarity or to avoid uncanny valley effects. Our research extends this literature on sign-language animation to explore **whether or not DHH signers prefer exaggeration in the movement of the character, specifically in regard to speed and timing parameters**. Analysis of participants' preferences in our study has revealed that some exaggeration of speed and timing can be beneficial.

Of course, there are many ways in which a movement may be exaggerated, and for creators of ASL animations, it is useful to know precisely **how to exaggerate in order to achieve greater user satisfaction**. Our findings have also addressed this issue, through collection of users' subjective evaluation of animations that vary across a range of speed and timing values for each parameter. Our work has revealed that DHH individuals with high/native fluency in ASL prefer for ASL animations to have slightly faster movements during signs, with slightly slower speed during transitions in-between words, and with less speed dynamics than would be typical in human signing.

In our examination of prior work, we had reported on one prior study [64] that had revealed a preference among DHH participants for ASL animations to be played somewhat slower than human signing, although that work had treated the speed of the ASL animation as a single variable. In our study, we investigated users' preferences in regard to five timing parameters that underlie modern sign-language animation synthesis technologies. Based on that prior study [64], we had expected for our participants to prefer values across all parameters that would result in slower-than-human animations. **However, our findings stand in partial contrast with that prior work [64]; since we investigated multiple timing parameters, our analysis revealed greater nuance.** To reconcile our findings with that earlier work, we speculate that the slower-than-human speed preferred by participants in that earlier study arose primarily from their preference for slower transitions and perhaps from the way that slower overall speed may mitigate any jumpy or unnatural animation during differential-rate speed-ups of signs—rather than from their preference for the speed of individual signs to increase or for there to be additional or longer prosodic pauses.

Multiple prior studies with DHH participants, e.g. [68, 72], have evaluated the quality of ASL animations along three key dimensions—understandability, naturalness of movement, and grammatical correctness—and consideration of these factors provides insight about why participants may have preferred timing parameters that differ from humans. Some preferences may be driven by **understandability**, especially as there are still limits in the quality of ASL animation technologies, which make them more difficult to understand. We speculate that displaying ASL with slower tran-

sitions in-between words may aid viewers in differentiating visually between spans of time when a sign is happening and spans of time when a transition is happening. In addition, displaying ASL with slightly faster sign duration could further emphasize this within-versus-between-signs distinction for the viewer. We speculate that as animation quality improves in the future, then users may begin to prefer animations to move at speeds closer to those of human signers. Preserving the **naturalness of movement** may be a factor in participants' preference for animations to have less extreme speed dynamics, as compared to human signers. Again due to limitations in the quality of current sign-language animation technology, animations may not maintain smooth movements or appearance when speeding-up or slowing-down individual signs, which we speculate is a reason why participants preferred for animations to be less dynamic than human signers. Similarly, naturalness may help explain why participants were happy with animations having similar pause length or pausing frequency as human signers, from the simple fact that the animated avatar does not need to move during pauses (aside from non-linguistic or physiologically-driven body movement, e.g. breathing). As compared to the challenges in producing natural movements during signs or transitions, it is relatively easier for creators to animate an ASL avatar to pause in a manner that looks natural—it just needs to stand still. Finally, preserving **grammatical correctness** may also help explain why participants preferred for ASL animations to pause in a manner similar to humans. We speculate that the linguistic meaning of pauses, e.g. to convey prosodic information in ASL or to suggest the syntactic structure of sentences, is so important to the message that it would be confusing if the frequency or duration of the pauses were to vary much from a typical human level.

More broadly, our findings suggest that other aspects of sign language animation should be investigated empirically with DHH users, rather than simply assuming that making animations as close as possible to humans in appearance and movement is the ideal. In this way, our work relates to a broader theme in the field of computing accessibility: **Designers and researchers should not assume that they know what is best** when designing for a specific group of users; there is value in substantial participation from users in the design and evaluation of technology intended for their use.

The outcome of this chapter address the fifth contribution of this dissertation:

Contribution 5: There is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation - for instance, prior work had found that users may prefer animations to be slower than human videos [63, 69]. Thus, for each of the timing parameters for

ASL animation, we empirically determined which values of that parameter are preferred by Deaf ASL signers via an experimental study, in which animations with a range of such values are displayed for comparison.

7.7 Conclusion

In this chapter, we conducted empirical studies to investigate speed and timing preferences among DHH ASL signers with high/native fluency, with aims: (a) to identify ASL signers' preferred values for these timing parameters and (b) determine whether participants prefer animations with timing values that differ from those in typical human signing. Specifically, our work investigated sign duration, transition time, differential signing rate, pause length, and pausing frequency. Our Pilot Study investigated the rang of preferred values for speed and timing parameters, Initial Five-Way Comparison Study identified the two most preferred values from among five values for each parameter, one being a typical human value, and a Final Two-Way Comparison Study identified the most preferred value.

While ASL signers preferred pausing-related parameters to be similar to human signers, they preferred ASL animations with faster-than-human sign durations, slower-than-human transition time, and with less extreme variation in differential signing speed. Our findings provide guidance for creators of future ASL animations or of animation-synthesis technologies, and our work demonstrates the importance of conducting studies with the participation of DHH ASL signers, rather than assuming these users will prefer animations to be as similar as possible to human signing.

Chapter 8

Investigating Acceleration Curves in ASL

The dissertation research presented thus far has considered speed during ASL as a singular multiplier for the hand movement within the ASL word or in-between words. Instead, the actual speed adjustments may, in fact, consist of acceleration profiles that affect particular sub-durations of each sign. This chapter investigates the nature of acceleration profiles during human recordings of ASL and how adjusting acceleration may improve the quality of ASL animations. The goal of this chapter is to address the final two contributions of this dissertation.

Contribution 6: Prior research on speed and timing of sign-language animation has not specifically investigated the issue of predicting acceleration curves for the movements of the character's body [7, 63, 69]. Further, some prior linguistic research has observed different classes of acceleration curves used during or between words in French Sign Language [36], but such an investigation has not been performed for ASL. Thus, we examine our new dataset to conduct an analysis of motion-capture patterns of human movements, to empirically determine whether there are common categories of acceleration curves present in different linguistic environments, e.g. within ASL signs, between ASL signs, or near sentence boundaries. This empirical

finding will inform the future design of acceleration curves for ASL-animation synthesis technology.

Contribution 7: Following the same logic as for Contribution 5 above, since there is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation, we empirically determine whether accuracy in the use of particular acceleration curves influences the subjective judgements of Deaf ASL signers, as to the quality of the ASL animation.

This chapter is structured in the following manner: section 8.1 will discuss prior related work on acceleration in other sign languages, e.g. Langue des Signes Française (LSF, French Sign Language). To investigate the role of acceleration in producing natural ASL animations, we examine recordings of human ASL signers section 8.2, with a goal to determine if there are specific classes of acceleration curves that appear during or in-between ASL signs. Next, in the Section 8.3 we report on a study with DHH users to evaluate whether animations that follow these acceleration patterns from human recordings are preferred to animations with simple linear acceleration curves.

8.1 Related Work

In research on human motion (originally within a sports context), Zatsiorsky discussed the movement primes that appear in a high-velocity movement [146]. Specifically, Zatsiorsky described two types of movements: The first type of movement are referred to as “ballistic movements," and they consist of movements during which the peak velocity occurs near the two-thirds point during the total time duration of the movement. Figure 8.1 illustrates the stereotypical velocity curve for a ballistic movement, and the red dot in Figure 8.1 represents the peak velocity. The second type of movement is referred to as a “controlled movement." As shown in Figure 8.2, the peak velocity during a controlled movement (see the red dot in Figure 8.2) usually occurs near the mid-point of the total time duration of the movement. As show in Figure 8.1 and Figure 8.2 a ballistic movement looks like a movement that has a "sudden stop" at the end. And in contrast, a controlled movement looks like a movement with a more consistent acceleration and deceleration.

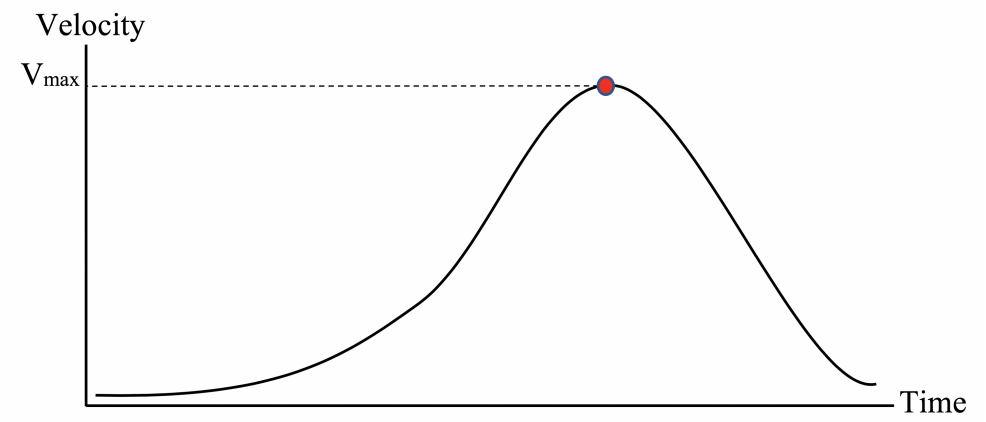


Figure 8.1: Velocity Curve for a Ballistic Movement

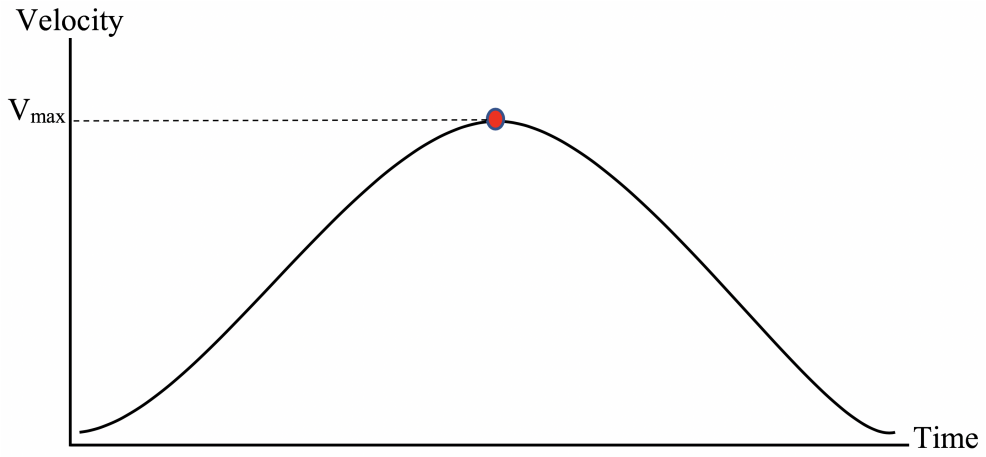
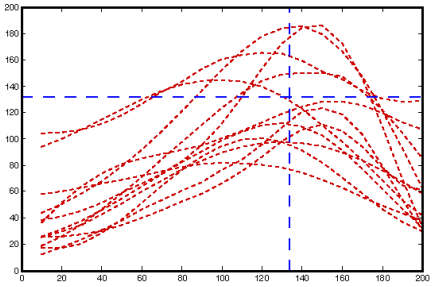


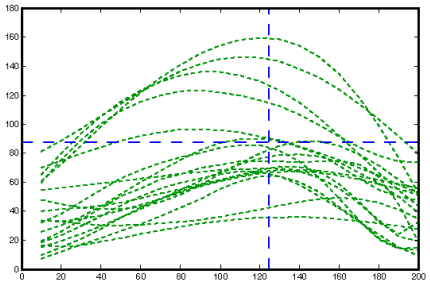
Figure 8.2: Velocity Curve for a Controlled Movement

Prior sign language research has used similar terminology; for instance, Johnson and Liddell have analyzed movement in ASL signs and found four types of movements: ballistic, fast ballistic, enduring, and extended enduring [79]. Yet, the fast-ballistic movement is simply a faster version of the ballistic movement, and the extended enduring movement is a slowed down version of the enduring movement. The enduring movements of Liddell and Johnson correspond to the controlled movements discussed previously.

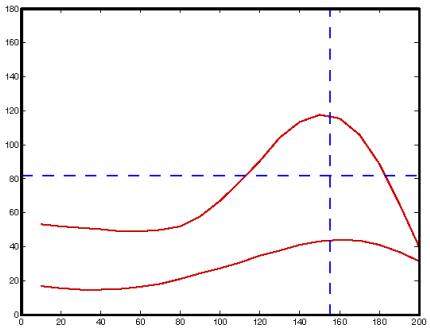
In fact, some recent motion-capture research on acceleration in French Sign Language has made use of a simplified taxonomy of accelerations, i.e. collapsing the distinction between ballistic and fast ballistic movements (Figure 8.1), and collapsing the distinction between enduring and extended enduring movements, which they refer to as controlled movement (Figure 8.2) [36]. In this previous work on French Sign Language [36], Kyle Duarte in his Ph.D. dissertation presented a detailed linguistic analysis in support of producing a realistic animated avatar in French Sign Language [36]. The author manually analyzed 22 data movements of French Sign Language motion-capture which was analyzed by human linguistic annotators. The annotators categorized segments of movements in the corpus into two motion curve types: ballistic (for ballistic and fast ballistic from Johnson and Liddell model [79]) and controlled (for enduring and extended enduring from Johnson and Liddell model [79]).



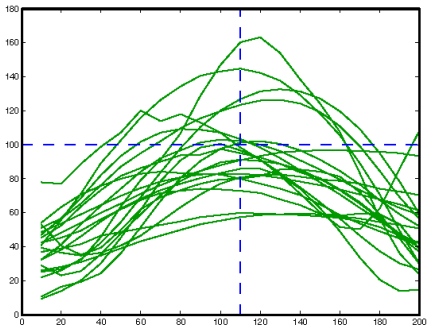
(a) Ballistic within-sign movements. Avg. V_{max} occurs at $.65 \times (t_x - t_0)$. $\sigma = .11$; $N = 12$.



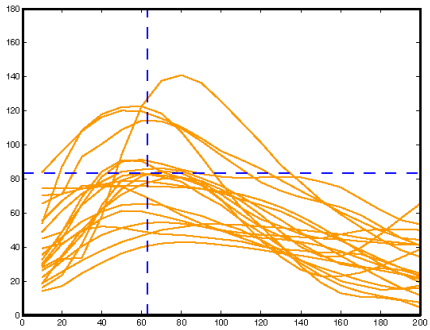
(b) Controlled within-sign movements. Avg. V_{max} occurs at $.60 \times (t_x - t_0)$. $\sigma = .12$; $N = 18$.



(c) Ballistic between-sign movements. Avg. V_{max} occurs at $.76 \times (t_x - t_0)$. $\sigma = .04$; $N = 2$.



(d) Controlled between-sign movements. Avg. V_{max} occurs at $.52 \times (t_x - t_0)$. $\sigma = .12$; $N = 20$.



(e) Inverse ballistic between-sign movements. Avg. V_{max} occurs at $.28 \times (t_x - t_0)$. $\sigma = .08$; $N = 20$.

Figure 8.3: These images showing velocity curves of hand movements during French Sign Language are reproduced from Duarte's dissertation [36]. Images (a-b) show movements within-signs, and (c-e) between signs. There were more examples of peak velocity occurring earlier for between-sign movements. The curves were standardized for time, so that they all had the same x duration (20 total x plot points); y values were left as-is for the purposes of maximum velocity comparison. The results are shown below.

Duarte analyzed these segments in two contexts: within-sign (movement curves that occur within sign boundaries) and between-sign (movement curves that occur between two sign). As shown in Figure 8.3, which extracted from Kyle Duarte Ph.D. theses [36], the analysis of within-sign movements found that there are 12 ballistic curves (with a peak velocity that occurred on average 65% of the way through the movement) and 18 controlled curves (with peak velocity that occurred on average 60% of the way through these movements) within LSF signs.

On the other hand, the results of between-sign movements found that there are three types of motion curve segments between signs, the first two motion curves are in ballistic and controlled movements (mentioned above), and the third motion curve between signs is inverse ballistic (which is the movement curves similar to ballistic curves but reflected, i.e. the peak velocity near the quarter of the movement). The author found two ballistic movements (with a peak velocity that occurred on average 76% of the way through the movement), 76 controlled movements (with a peak velocity that occurred on average 52% of the way through the movement), and 20 inverse ballistic movements (with a peak velocity that occurred on average 28% of the way through the movement). The result of between-sign movements analysis suggests that there are three groups of curves between signs in French Sign Language, and overall there are more examples of curves with peak velocity occurring sooner during between-sign contexts. However, there has been no previous work that has investigated this issue in ASL, nor for the purpose of generating ASL animation.

8.2 Modeling Acceleration Curves in ASL

Through an analysis of motion-capture patterns of human movements in our corpus of ASL multi-sentence passages (chapter 4), the goal of this study is to determine whether there are common varieties of acceleration curves present in different linguistic environments, e.g. within ASL signs, between ASL signs, or near sentence boundaries.

8.2.1 Research Question

The study presented in this section addresses the following research question:

- What is the distribution of acceleration curves in human ASL signing?

Secondarily, we also consider whether this distribution appears similar to that found in within-sign and between-sign contexts in LSF.

8.2.2 Method

To investigate acceleration curves in ASL, we employ a two-stage methodology: First, we extracted useful motion velocity information from our ASL motion-capture dataset (Sub-subsection 8.2.2.1 Dataset Engineering), and then we explored the acceleration patterns in this data (in Sub-subsection 8.2.2.2 Results for Velocity Distribution). The details of this methodology are explained in the following sections.

8.2.2.1 Dataset Engineering

We first processed and extracted relevant information from our ELAN ASL dataset, in a similar manner to how data processing had occurred in earlier chapters, which had also made use of both the motion-capture data of hand movements and the accompanying linguistic annotation data. For the current study, we specifically extracted the velocity of the signer's hand so that we could partition the entire timeline into three types of sub-durations: *movement sections* during signs, *movement sections* between signs, and *hold sections* (periods in which there is not movement) during signs, along the lines of the Movement-Hold model of (Liddell and Johnson, 1989) [95]. This model is a linguistic representation of the time structure of ASL signs, and a sample representation of this model is shown in Figure 2.1. The outcome of this data process was a flat CSV file, for four ASL signers, with timing values and annotation information (movements or holds). For analysis of this velocity data, our main focus is on the *movement sections*, which are considered separately in two contexts, each of which is stored in a separate data file during our analysis:

- Movement sections occurring internally during an ASL word
- Movement sections occurring between ASL words

8.2.2.2 Results for Velocity Distribution

The Tableau[18] statistical analysis software was used to visualize the data and perform data exploration, as discussed below.

[18] https://www.tableau.com/

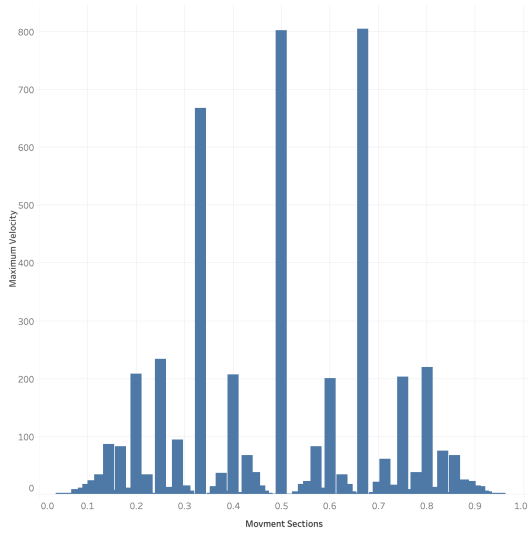


Figure 8.4: Distribution of the time (normalized on a 0 to 1 scale) when the maximum velocity occurred during each movement section within ASL signs.

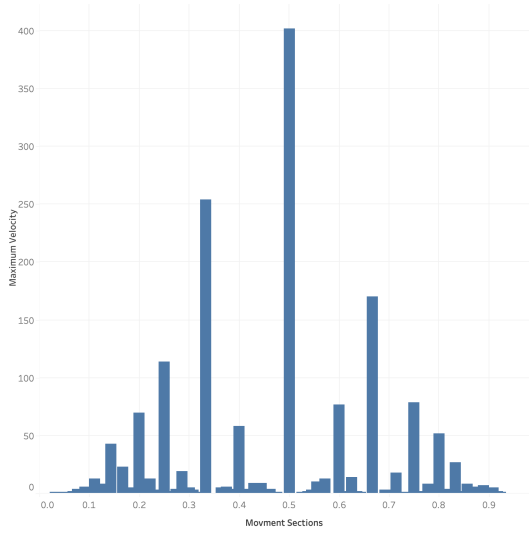


Figure 8.5: Distribution of the time (normalized on a 0 to 1 scale) when the maximum velocity occured during each movement section between ASL words (but excluding any between-sign sections at sentence boundaries).

Figure 8.4 illustrates the distribution of the time when maximum velocity occurred during the timeline of every movement section (normalized from 0.0 to 1.0) within ASL words. Our distribution analysis revealed that the maximum velocity generally occurred during a range during the middle third of the timeline during each movement, with the peak slightly shifted toward the second half of each movement.

Figure 8.5 illustrates the distribution of the time when maximum velocity occurred during the timeline of every movement section (normalized from 0.0 to 1.0) between ASL words. (Movements that occur in-between words that appear at sentence boundaries were excluded from this analysis, as prior work in this dissertation had suggested that there may be unique timing characteristics at sentence boundaries.) As compared to the distribution for within-word movements, we noted that there was a tighter concentration of velocity peaks near the mid-point of the duration of each movement, and there were also more examples of movements in which the maximum velocity occurred during the first half of the movement duration.

8.2.2.3 Comparison to Prior LSF Research

In Duarte's prior analysis of LSF, 2 types of acceleration curves had been observed during within-sign movements, i.e., curves with maximum velocity during the middle and curves with the maximum velocity near the two-thirds point on the duration, and 3 types of curves had been observed during between-sign movements, i.e., with maximum velocity during the first half, during the middle, or during the second half. To aid the reader in comparing these prior results for LSF with our new results for ASL, two composite figures have been produced: figure 8.6 and figure 8.7.

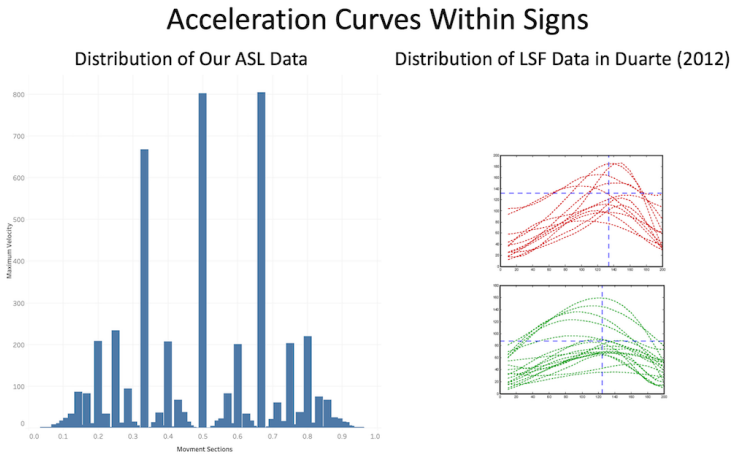


Figure 8.6: Composite image of our previously shown results for within-sign movements in ASL, as compared to previously shown within-sign curves for LSF.

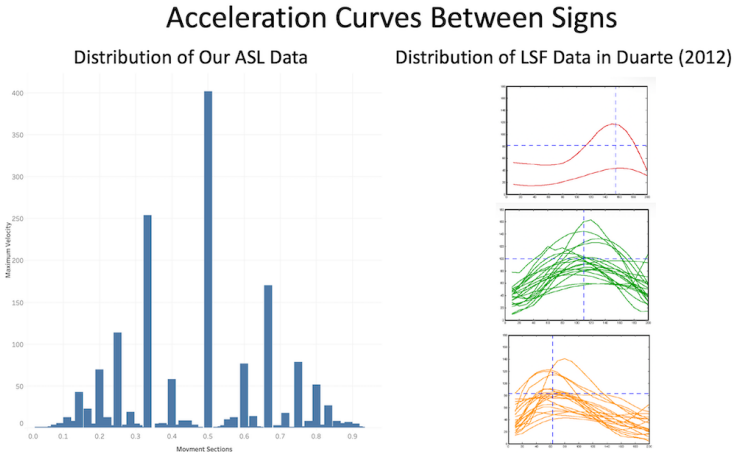


Figure 8.7: Composite image of our previously shown results for between-sign movements in ASL, as compared to previously shown within-sign curves for LSF.

For instance, figure 8.6 contains on the left side our distribution graph showing within-sign maximum-velocity times for ASL, and on the right side, the two graphs of velocity for within-sign movements for LSF are shown from Duarte's dissertation. If the reader were to envision merging these two LSF distributions, then they could be compared to the overall distribution data shown for ASL. Similarly, if the reader were to envision merging the three sub-graphs on the right side of

figure 8.6, the reader may compare this to the ASL data for between-sign locations for ASL shown on the left. Overall, we observe that the distributions between sign in ASL and LSF have a relatively similar shape, i.e., as shown in figure 8.7, if we merge the three right images for LSF, they would correspond to the slightly left-shifted distribution of the data shown in the large ASL graph on the left side of figure 8.7).

While the Duarte dataset is too small to support rigorous statistical comparisons between ASL and LSF maximum-velocity timing during movements, our findings are suggestive of there being unique distributions during within-sign and between-sign movements during ASL, along with the overall shape of these distributions being loosely consistent with those those that had been observed in prior work on LSF [36].

8.3 User Study

For ASL animation researchers, the analysis presented in the previous section suggests that when planning acceleration curves for movements during ASL, there may be a benefit in making use of individual acceleration distributions for within-sign movements and for between-sign movements. Similar to earlier phases of this dissertation research, it was important for us to conduct a study with users to determine whether animations based on acceleration curves from human ASL movements would lead to better ASL animations. We therefore conducted a study with DHH participants to evaluate whether participants would prefer ASL animations in which the acceleration curves were drawn from the distributions in this human motion recordings, or whether they prefer animations with uniform distributions. Notably, animations with uniform distributions, i.e., with the peak velocity always at the mid-point of any movement, are the default in existing ASL animation software, such as Sign Smith Studio.

8.3.1 Research Question

This section addresses the following research question:

- Do DHH users prefer animations that follow the acceleration-curve distributions observed in motion-capture recordings of human ASL signers, or animations that follow uniform accelerations curves that are currently typical in ASL animation software?

8.3.2 Method

As in our prior user studies, we created a new set of ASL animations using the Sign Smith Studio animation software. Then, we asked a researcher from our lab who is a fourth-year ASL interpreting student who had taken courses on ASL linguistics and had several years of experience working as an ASL linguistic analysts performing video-analysis of ASL recordings for our team to mark the movement and hold segments of the words in those animations. (Note: The researcher in this case was analyzing an animation, not a video of a human; the reason for this unusual annotation tasks is that we needed to post-process the animations to manipulate them.) In addition, the researcher labeled which of the movements occurred during words or in-between words.

To produce the animation stimuli for our study, we began with a very high-frame-rate version of each animation (specifically, four-times more frequent frames per second, as compared to typical 30 frames per second ASL animations). Using the linguistic annotation timeline of all of the movement sections (produced by the researcher above), we were able to insert and remove frames from this source animation file, in order to produce a desired animation, in which the maximum velocity of each movement occurred at a particular point during the duration of the movement.

During this process, our aim was to produce a resulting 30 frames-per-second version of each animation, in which the maximum velocity of the movement occurred during a particular moment during the timeline of each individual movement. To decide where we should set the maximum velocity moment during the duration of each movement, we randomly sampled from the dataset we had collected from human ASL signers, as presented in the prior section. For within-sign movements, we randomly selected a curve from the dataset of within-sign movements, and for between-sign movements, we randomly selected a curve from the dataset of between-sign movements. Thus, for example, if we selected a curve with maximum velocity occurring at the 0.6 point during the 0-to-1 timeline duration of the movement, then we deleted or inserted frames from our source animation file so that the maximum velocity of that individual movement occurred at the 0.6 point during the duration of the final animation.

We used Sign Smith Studio [133] to generate animation stimuli because we found that software enabled easy editing of timing parameters when producing ASL animations. We used the same three ASL passages used in that have been used in the previous studies [10], with each passage approximately 75 words in length and discussing following topics: a bear, the cost of rice, and a short biography of a university student.

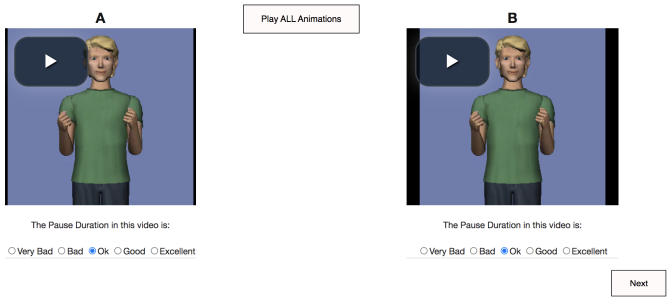


Figure 8.8: A sample screenshot for one of the pairs of animations displayed in the user study.

For each original ASL animation, we produced two versions: The first version corresponded to default acceleration settings of the Sign Smith Studio software, in which the maximum velocity of each movement occurred at its midpoint. The second version corresponded to an animation in which the acceleration curve for each movement was based on a randomly selected example from either our within-sign dataset or our between-sign dataset, as appropriate. Figure 8.8 illustrates the side-by-side animations of virtual human character performing ASL for the original a modified animations.

We conducted a user study by recruiting 13 participants who were fluent ASL signers, and they were asked to share their feedback about the new animations (with modified acceleration) compared to a baseline (with uniform acceleration). While we had used in-person studies in prior chapters of this dissertation, we conducted this study in an online remote manner, due to the COVID-19 pandemic.

8.3.2.1 Participants

At the beginning of the appointment, the participants answered demographic questions. Our 13 participants included 10 who self-reported as “female" and 3 as “male." Their ages ranged from 20 to 52 years old, with a median age of 29. Eleven participants identified as Deaf, one hard-of-hearing, and one hearing. Twelve participants learned ASL since early childhood. Eleven participants used ASL at home, and 12 used ASL as a young child at school. All participants reported using ASL at college or university.

8.3.2.2 Results

Figure 8.9 shows the average subjective response score for the original animations (generated using default acceleration settings for Sign Smith Studio software, shown in orange color on the figure) and for the new animations (generated using the new distribution of acceleration curves based on human ASL recordings, blue color). Higher score in figure 8.9 means more preferred animations, and the participants slightly preferred the new animation with modified speeding values.

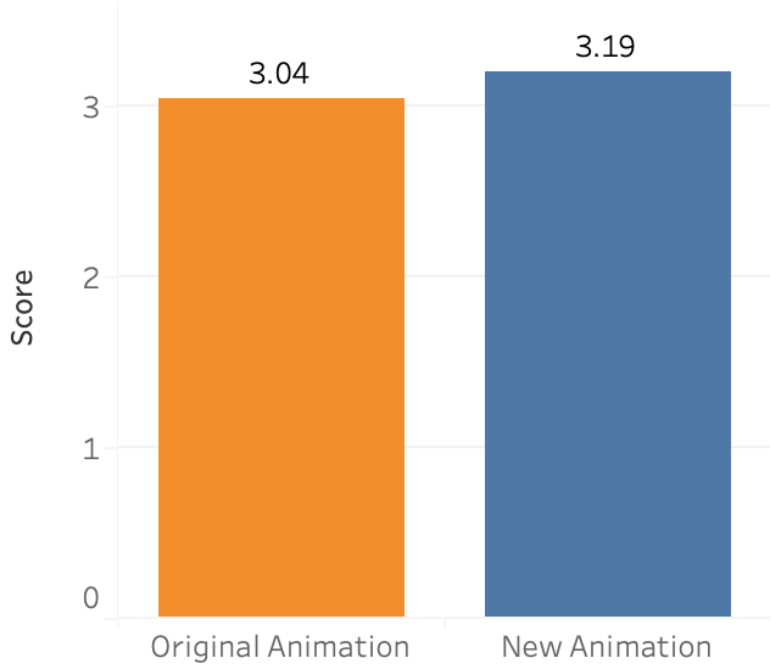


Figure 8.9: Participants' average subjective preference scores for the baseline and the new animations shown in the study.

We performed a paired t-test on the subjective response scores from participants in the study, which revealed no significant difference between the average subjective scores for participants for the two versions of animation: the original animation and the animations with modified speed profile (p value = 0.425).

Participants were not told how the two animations differed, nor that they differed at all. After viewing each pair of animation, participants were asked "Is there any difference between the animations?" Participants' responses indicated that they noticed that the message of the animations

was the same, but their responses indicated that they did notice that there was a difference in speed between the two animations, e.g. with some participants responding:

- *Aside from the speeding, nothing.* [P03]
- *No, I don't see any differences between A and B aside from the speed.* [P04]
- *I feel like they're the same aside from the speeding.* [P05]
- *Nope, other than speed.* [P12]

During an interview question after each of the 2 pairs of animations shown, each participant was asked which of the two animations they preferred. A majority of participants indicated that they preferred the new animation with modified speed profile: 14 out of 26 video were preferred as the new animation, 11 original animation were preferred, and 1 participant said that the two animations shown appeared to be the same. These responses further suggest that participants noticed that there was some difference between the two animations (except in the case of the one pair of animations which one participant believed appeared to be the same). Despite the difference in the movements of these animations being relatively subtle, this finding suggests that fluent ASL signers who view animations with manipulations to the acceleration curves of specific movements are sensitive to this aspect of animation speed.

Among those participants who preferred the "NEW" animations with acceleration curves based on human motion, participants shared open-ended comments explaining their preferences:

1. *I like [NEW], as I could notice things better, their signing seems less fast than the other. Like, [OLD] is faster (not too fast but just a bit), you know.* [P05]
2. *I picked [NEW] this time because it's more natural and clear.* [P05]
3. *I would prefer [NEW], Because it's normal paced.* [P09]
4. *[OLD] is too fast. I could understand most signing but the spelling was too fast that it was distorted, so I couldn't understand it.* [P09]
5. *I prefer NEW because it's slower and gives me time to adapt to its style.* [P04]
6. *[OLD] needs a lot of improvement. [OLD] is too fast, nothing about [NEW], it's just perfect.* [P08]

7. *Honestly, I would like [NEW] better. I couldn't see the spelling in [OLD], but [NEW] is smooth and it matches the spelling, so yes, the speeding is important for spelling.* [P11]

When providing comments about the animations shown in the study, participants pointed to the specific animations they were referring to, or referred to them as the "left" or "right" animation. The left-or-right arrangement of animations was randomized across the study, and when transcribing the comments above, references have been replaced with "[OLD]" or "[NEW]," to refer to animations with uniform accelerations or animations with accelerations based on the distribution in human recordings, respectively.

Of course, some participants preferred the "[OLD]" animation in each pair, and they offered a variety of reasons for this preference, often focusing on their perception that the "[NEW]" animations appeared faster:

1. *Just a little too fast. When they were signing, it's okay but when they move their hands out, they did that too fast; it should be at the same speed as the signing.* [P07]

2. *[OLD] is better than [NEW] because [NEW] is more stiff, they're the same speed but [OLD] is soft while [NEW] has sudden lag/movements.* [P10]

3. *The speed is a bit slower, and has more body language, but [NEW] is at the speed that you're not sure about their facial expressions. [OLD] is not at the best speed but I can notice more while [NEW] just throws it out there.* [P12]

It is important to note that the overall words-per-minute of the two versions of each animation were identical, aside from the subtle difference in the acceleration curves for the movements. The overall duration of each mo vement was identical, and if the two animations were played simultaneously, then on a word-by-word basis, the two animations would appear to be synchronized. So, the perception by some participants that the "[NEW]" animations seemed faster appeared to be based on the difference in acceleration curves during the movements.

8.4 Conclusion

The studies presented in this chapter have addressed the sixth and the seventh contribution of this dissertation:

Contribution 6: Prior research on speed and timing of sign-language animation has not specifically investigated the issue of predicting acceleration curves for the movements of the character's body [7, 63, 69]. Further, some prior linguistic research has observed different classes of acceleration curves used during or between words in French Sign Language [36], but such an investigation has not been performed for ASL. Thus, we examine our new dataset to conduct an analysis of motion-capture patterns of human movements, to empirically determine whether there are common categories of acceleration curves present in different linguistic environments, e.g. within ASL signs, between ASL signs, or near sentence boundaries. This empirical finding will inform the future design of acceleration curves for ASL-animation synthesis technology.

Contribution 7: Following the same logic as for Contribution 5 above, since there is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation, we empirically determined whether accuracy in the use of particular acceleration curves influences the subjective judgements of Deaf ASL signers, as to the quality of the ASL animation.

Specifically, our studies have revealed that human ASL signing includes a range of acceleration curves during movements, and our analysis was suggestive that there may be differences in the distribution of these curves during within-sign and between-sign contexts. An informal comparison to prior work on LSF revealed that our findings for ASL aligned with the distribution of acceleration curves observed in LSF, for within-sign and between-sign contexts. Finally, our user study with fluent ASL signers revealed that participants were able to notice a difference between animations that differed only in this subtle aspect of movement (the acceleration curve for each movement), with all other speed and timing aspects of the animation remaining identical (including the overall video length). While we report average subjective preferences from participants in the study, figure 8.9), the small sample size of this study did not reveal statistically significant differences in viewers' subjective preferences between animations based on uniform acceleration curves and those based on a distribution of acceleration curves from human recordings.

EPILOGUE FOR PART III

In part III we presented our empirical evaluation of participants' preferences for the speed and timing parameters in the ASL animations. In Chapter 7 we conducted several empirical studies to investigate speed and timing preferences among DHH ASL signers with high/native fluency, with two aims: (a) to identify ASL signers' preferred values for these timing parameters and (b) determine whether participants prefer animations with timing values that differ from those in typical human signing. We found that while ASL signers preferred pause length and frequency to be similar to those of humans, they actually preferred animations to have faster signs, slower transitions, and less dynamic variation in differential signing speed, as compared to the timing of human signers. Our findings provide specific empirical guidance for creators of future ASL animation technologies, and more broadly, it demonstrates that it is not safe to assume that ASL signers will simply prefer for properties of ASL animations to be as similar as possible to human signers.

Chapter 8 focused on acceleration curves in ASL during movements in ASL signs. Beginning with a summary of prior work on acceleration curves in LSF sign movements, we then conducted an analysis of hand velocity data from recordings of ASL human signers from our motion-capture corpus. This analysis revealed distributions of when the maximum velocity occurs during movements in within-sign and between-sign contexts in ASL, and the results were suggestive of similarity to the distributions observed in prior work on LSF. Finally, we conducted a user-based evaluation with fluent ASL signer to investigate viewers' preferences between animations with simplistic uniform acceleration during movements, as compared to animations with acceleration curves based on human recordings.

Chapter 9

Conclusions and Future Work

This chapter begins with a high-level overview of the research activities of this dissertation research, as well as information about the findings or achievements from these activities. Next, a brief summary of the main contributions of this dissertation is provided, and finally, this chapter includes a set of potential avenues of future work that could be investigated beyond this dissertation.

9.1 Summary of Research Activities

This dissertation has examined the use of a data-driven approach to model timing parameters to generate realistic ASL animations. We used motion-capture data recorded from humans to train machine learning models to predict realistic timing parameters for ASL animation, with a focus on inserting prosodic breaks (pauses), adjusting the pause durations for these pauses, and adjusting differential signing rate for ASL animations, based on the sentence syntax and other features, which we had engineered and extracted, and which were inspired by prior linguistics literature on ASL. We evaluated our data-driven approach using different levels of evaluation: starting with selecting a robust model using cross-validation, then comparing our modeling with the rule-based approach. In all of those evaluations, our models beat the baseline for comparison. Then, we conducted multiple user studies with DHH participants to gather human feedback about the generated animations. We also conducted studies to investigate participants' preferences about the best values for speed and timing in animation and whether these values should be similar to human values. Finally, we

examined the distribution of animation curves in ASL and how their distribution compared to those in prior work on another sign language. Across all of this work, our main goal was to create new knowledge that would inform how to best automate the selection of speed and timing parameter for ASL animation synthesis, so that the the resulting ASL output is understandable.

In Part I, we discussed the process of adding layers of annotation atop our prior motion-capture corpus, in order to support speed and timing research (Chapter 4). In addition, we discussed transferring the original release of motion-capture corpus to the ELAN platform, and documented our workflow for preparing the dataset for speed and timing modeling research. This dataset produced during Part I is then used throughout later stages of the dissertation work.

In Part II, we discussed our methodology for speed and timing modeling and our evaluation of this work. We explained the process of building machine-learning models, selecting the best subset of features, and tuning model parameters to obtain a robust model via cross-validation. We evaluated our models by comparing them with a baseline set of typical ASL speed and timing values used in simple ASL animations, and we examined the quality of our best model, by comparing it with a prior state-of-the-art rule-based approach. We found that our system out-performed this prior model in predicting timing values for a set of ASL passages (Chapter 5). Then, we generated ASL animations (using our model and using a baseline model), and we conducted an interview with DHH participants to understand their preferences and feedback about the resulting ASL animations. We found that participants preferred our animations, and the majority of participants offered positive comments about the animation generated using our models of speed and timing in ASL animations (Chapter 6). Given the importance of evaluating systems in studies with real users, and the promising results of our interview with DHH participants, we decided to conduct a larger experimental study with Deaf native ASL signers; this larger study is a focus of Part III.

In Part III, we investigated whether DHH signers prefer timing values in animations that are identical to those in human recordings, or whether they perhaps prefer faster, slower, or more exaggerated values. We investigated five speed and timing parameters: sign duration, transition time, differential signing rate, pause length, and pausing frequency. We first conduct a pilot study with users to identify an appropriate range of speed and timing parameters [9] to focus on in later studies, and then we conducted an *Initial Five-Way Comparison Study* with users. This study revealed the two most preferred values for each speed and timing parameter (one of which was a typical human value for this parameter). Finally, we conducted our *Final Two-Way Comparison Study* to determine whether participants prefer animations with timing values that differ from those in typical

human signing. We found that, while ASL signers preferred pausing-related parameters to be similar human signers, they preferred ASL animations with faster-than-human sign durations, slower-than-human transition time, and with less extreme variation in differential signing speed. Finally, our work in Part III focused on investigating acceleration curves in ASL. We analysed the distribution of these human recordings of ASL signing, compared those distributions to those in another sign language, and conducted a user study to determine whether users prefer for acceleration curves in ASL animations to be based on those of human recordings.

9.2 Contributions

The key contributions of this dissertation are presented in the following paragraphs.

Contribution 1: We **created a new American Sign Language Speed and Timing Dataset**, which was an enhancement to our lab's pre-existing motion-capture corpus of ASL. As part of this work, we transferred an existing motion-capture corpus to a new linguistic annotation platform that has become standard among sign-language linguistic researchers, ELAN [130]. This work included adding layers of linguistic annotation and documenting our data pre-processing procedures. The transformations and enhancements to this corpus were necessary to make this resource useful for subsequent feature engineering and to provide input for our machine-learning modeling work. We released this new dataset to the research community so that future researchers can use it in their work.

Contribution 1 was discussed in Chapter 4.

Contribution 2: **We empirically determined which set of features was most influential in our speed and timing prediction models**, via a feature-ablation analysis. Since the motivating goal of our research is to enable future animation systems to convert a script that specifies an ASL message into an animation automatically, we identified the minimal set of information that the person writing the script must specify in order for our software to operate. We focused on the features that are useful for our three modeling tasks:

2.A: We empirically determined the best subset of features needed to be used for building a predictive model of the **prosodic breaks (pauses)** after each word. The important features for this model included: Sentence_Boundaries, Clause_Boundaries, Noun_Phrase_Boundaries, Verb_Phrase_Boundaries, Sentence_Length, Noun_Phrase_Length, Verb_Phrase_Length, Relative_Proximity, and Complexity_Idx.

2.B: We empirically determined the best subset of features needed to be used for prediction the **time-duration of each break or pause**. In addition to the features from contribution 2.A, we also used the following features: Pausing, Pausing_Before_Gloss, Word_Order_On_Sentence, Reverse_Word_Order_On_Sentence, Word_Duration, and Next_Word_Duration.

2.C: We determined the best subset of features that should be used for modeling the **variation of the speed of each particular word** in the message. In addition to the features used in 2.A, the following features: Pausing, Pausing_Before_Gloss, Word_Order_On_Sentence, and Reverse_Word_Order_On_Sentence features, were all useful for this differential rate modeling.

Contribution 2 was discussed in Chapter 5.

Contribution 3: We empirically determined that a machine learning modeling trained on the final subset of the linguistic features **out-performed prior state-of-the-art rule-based** approaches for the task of predicting the timing parameters for ASL multi-sentence passages, based on a cross-validation analysis of held-out data. Models were created and evaluated for all of the following speed and timing parameters:

3.A: Whether a **prosodic break (a pause)** should occur after each specific word.

3.B: What the value of the **time-duration of any such break/pause** should be.

3.C: What the **variation in speed (slightly faster, slightly slower) should be for each particular word** in the message.

Contribution 3 was discussed in Chapter 5.

Contribution 4: In a user-based study, we found that **DHH ASL signers preferred animations of multi-sentence ASL passages in which timing values were determined by our new models**, which we had investigated in contribution 3, rather than when these timing values had been determined by the a rule-based technique.

Contribution 4 was discussed in Chapter 6.

Contribution 5: There is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation. For instance, prior work had found that users may prefer animations to be slower than human videos [63, 69]. Thus, **for each of the timing parameters for ASL animation, we empirically determined which values of that parameter are preferred by Deaf ASL signers via an experimental study**, in which animations with a range of such values are displayed for comparison.

Contribution 5 was discussed in Chapter 7.

Contribution 6: Prior research on speed and timing of sign-language animation had not specifically investigated the issue of predicting acceleration curves for the movements of the character's body [7, 63, 69]. Further, some prior linguistic research has found different classes of acceleration curves used in intra- and inter-word contexts in French Sign Language [36], but such an investigation has not been performed for ASL. Thus, we conducted an analysis on our new dataset of motion-capture patterns of human movements, **we empirically determined whether there are common categories of acceleration curves present in various contexts**, e.g. within ASL signs, or between ASL signs. This empirical finding will inform the future design of acceleration curves for ASL-animation synthesis technology.

Contribution 6 was discussed in Chapter 8.

Contribution 7: Following the same logic as for Contribution 5 above, since there is a possibility that the range of timing values used by humans could differ from the range of timing values DHH ASL signers would like to see in an animation, **we empirically investigated whether animations that contain acceleration curves that are**

based on our analysis of recordings of human ASL signers would receive different subjective judgements of Deaf ASL signers, as compared to animations using baseline linear accelerations common in current sign-language animation systems. While analysis of the subjective response scores in our study did not reveal a statistically significant difference, participants' open-ended responses indicated that fluent ASL signers were able to notice the difference between animations that differed only in this subtle aspect of movement.

Contribution 7 was discussed in Chapter 8.

9.3 Limitations and Future Work

In Chapter 4, we discussed our approach to **extending our motion-capture dataset** by adding additional annotation for additional ASL signers; however, the size of our generated dataset is still relatively small. Future research could collect and annotate additional ASL motion-capture data in order to investigate improving the machine learning modeling task.

Other limitations related to **how stimuli were created, displayed and measured** when evaluating speed and timing models for ASL animations. The user studies in this work have made use of a short set of ASL passages as the basis for the stimuli in our experimental studies, and future work should investigate speed and timing parameters using ASL passages with a wider variety of topics, genres, and lengths.

Another limitation related to the number of **participants** in studies presented throughout this dissertation. While we endeavored to conduct studies with as large a number and as diverse a group of participants as possible, most of our user studies included a relatively small number of participants and their composition do not fully reflect the entire Deaf community. For example, in Chapter 7, we included 56 DHH participants across all of our studies, and while this is a relatively large number of participants for research in the field of ASL animation synthesis, the participants in our study certainly did not reflect the full diversity of all ASL signers. For instance, although we recruited through online advertisements, the majority of our participants were university students who were relatively young. In future work, there is a need to conduct studies with a larger number of participants, across multiple dimensions of diversity, to ensure that our findings generalize to those groups.

One such dimension is the *level of ASL experience* and skill of the participant. In our study, we recruited specifically for participants with a large amount of ASL experience and fluency, i.e., those who began using ASL since infancy or early childhood. However, there are many users of ASL who learned the language later in life or who are still learning or developing their language skills, and it would be important to determine whether the speed and timing preferences of those users may differ from those of the participants in our study.

Another limitation of our work is that, in Chapter 7, we have investigated levels of each *parameter individually* (using default values for the other parameters while we investigated users' preferences for each parameter). A future study would be needed to investigate interaction effects between these parameters.

Furthermore, a possible limitation of the Initial Five-Way Comparison Study and Final Two-Way Comparison Study user studies that we conducted in Chapter 7: Since those studies were conducted remotely due to COVID, there is a risk that poor internet connection for the user could have affected their video. Thus, future work could repeat this study in an in-person modality.

Chapter 8 presented findings about the shape of the distribution of acceleration curves in ASL. Future work is needed to investigate whether there are differences between the distribution of the acceleration curves not only within-signs and between-signs, but also for any particular subset of movements in ASL. Further characterization or classification of those types of curves may inform future work on selecting appropriate acceleration curves for ASL animations. Chapter 8 also investigated whether fluent ASL signers noticed differences between animations that differed only in regard to their acceleration curves. However, the number of participants in that study was relatively small (13 participants). Future work with a larger number of participants may be needed in order to investigate this issue empirically.

The research in this dissertation depends upon the dataset used for model training. In this dissertation we built variety of models to predict the speed and timing parameters for ASL, future work could use the modeling approaches in this dissertation for other sign languages, assuming that future researchers were able to provide a corpus with the same type of linguistic annotations. Furthermore, if future researchers were to collect a larger dataset of ASL, then it would be interesting to re-evaluate the modeling of speed and timing values using a leave-one-signer-out cross-validation approach, in which the signer whose data is in the testing set does not appear in the training set, for each "fold" of the cross validation. Another possible avenue to consider for future datasets would involve collecting a dataset with more detailed or richer annotations, e.g., that include more

linguistic details about each word or its phonological details. With such a dataset, future work could examine additional features or try other types of modeling, in order to predict speed and timing values for ASL or other sign languages. Finally, future work might use our existing speed-and-timing ASL dataset for other research, e.g., for linguistic research into sign-language coarticulation effects, on the speed of finger movements in handshape change, or the speed of wrist movements for palm-orientation change.

9.4 Conclusion

This dissertation has investigated American Sign Language (ASL), which is a primary means of communication for over half million people in the U.S. A key motivation for this research has been that many people who are Deaf or Hard of Hearing (DHH) prefer to receive information in the form of ASL. Unfortunately, few websites present their information content in the form of ASL. A challenge is that videos of human ASL signers would be difficult to update and maintain when information on a website must change. This dissertation has investigated technology to automate the creation of animations of ASL based on easy-to-update script. This dissertation has provided specific guidance for speed and timing in ASL animations, for creators of future ASL animation, which may include artists who are animating virtual humans or researchers building models of speed and timing for animation synthesis. This dissertation has demonstrated a mix of data-driven and user-based research, which may provide an example for how other aspects of sign-language animation could be investigated, e.g., for other signed languages or for other linguistic aspects beyond speed and timing. The goal of this research is to automate this aspect of animation synthesis and to create understandable and realistic ASL animation with minimum human effort. Our hope is that advancements in ASL animation synthesis technology will support access to information for DHH users in many new contexts.

Bibliography

[1] Nicoletta Adamo-Villani and Saikiran Anasingaraju. 2016. Toward the Ideal Signing Avatar. *EAI Endorsed Transactions on e-Learning* 3, 11 (2016). (Cited on pages 95, 120).

[2] Nicoletta Adamo-Villani and Ronnie B Wilbur. 2015a. ASL-Pro: American Sign Language Animation with Prosodic Elements. In *Universal Access in Human-Computer Interaction. Access to Interaction*. Springer International Publishing, Cham, 307–318. DOI:10.1007/978-3-319-20681-3$_2$9 (Cited on pages 2, 14, 17, 23, 28, 30, 31, 33, 35, 36).

[3] Nicoletta Adamo-Villani and Ronnie B Wilbur. 2015b. ASL-Pro: American Sign Language Animation with Prosodic Elements. In *Universal Access in Human-Computer Interaction. Access to Interaction*. Springer International Publishing, Cham, 307–318. DOI:10.1007/978-3-319-20681-3$_2$9 (Cited on page 119).

[4] Jake K Aggarwal and Quin Cai. 1999. Human motion analysis: A review. *Computer vision and image understanding* 73, 3 (1999), 428–440. (Cited on page 13).

[5] Hend S Al-Khalifa. 2010. Introducing Arabic sign language for mobile phones. In *International Conference on Computers for Handicapped Persons*. Springer, 213–220. (Cited on pages 95, 120).

[6] Sedeeq Al-khazraji. 2018. Using Data-Driven Approach for Modeling Timing Parameters of American Sign Language. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 497–500. (Cited on page 62).

[7] Sedeeq Al-khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2018a. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 259–270. DOI:10.1145/3234695.3236356 (Cited on pages 7, 16, 24, 62, 80, 91, 98, 99, 123, 139, 145).

[8] Sedeeq Al-khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2018b. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 259–270. DOI:10.1145/3234695.3236356 (Cited on pages 95, 119).

[9] Sedeeq Al-khazraji, Becca Dingman, and Matt Huenerfauth. 2020a. Empirical Investigation of Users' Preferred Timing Parameters for American Sign Language Animations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7. (Cited on pages 16, 98, 142).

[10] Sedeeq Al-khazraji, Becca Dingman, and Matt Huenerfauth. 2020b. Empirical Investigation of Users' Preferred Timing Parameters for American Sign Language Animations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7. (Cited on pages 94, 95, 97, 109, 111, 112, 115, 134).

[11] Sedeeq Al-khazraji, Becca Dingman, Sooyeon Lee, and Matt Huenerfauth. 2021. At a Different Pace: Evaluating Whether Users Prefer Timing Parameters in American Sign Language Animations to Differ from Human Signers' Timing. (Cited on page 16).

[12] Sedeeq Al-khazraji, Sushant Kafle, and Matt Huenerfauth. 2018. Modeling the Use of Space for Pointing in American Sign Language Animation. In *In Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, The 11th International Conference on Language Resources and Evaluation (LREC 2018)*. (Cited on pages 16, 62).

[13] Abdulaziz Almohimeed, Mike Wald, and Robert Damper. 2010. An Arabic Sign Language corpus for instructional language in school. (2010). (Cited on page 4).

[14] Sylvain Arlot, Alain Celisse, and others. 2010. A survey of cross-validation procedures for model selection. *Statistics surveys* 4 (2010), 40–79. (Cited on page 38).

[15] Norman I Badler, Martha S Palmer, and Rama Bindiganavale. 1999. Animation control for real-time virtual humans. *Commun. ACM* 42, 8 (1999), 64–73. (Cited on page 13).

[16] Norman I Badler and Stephen W Smoliar. 1979. Digital representations of human movement. *ACM Computing Surveys (CSUR)* 11, 1 (1979), 19–38. (Cited on page 13).

[17] JA Bangham, SJ Cox, Michael Lincoln, I Marshall, M Tutt, and Mark Wells. 2000b. Signing for the deaf using virtual humans. (2000). (Cited on page 28).

[18] J Andrew Bangham, SJ Cox, Ralph Elliott, JRW Glauert, Ian Marshall, Sanja Rankov, and Mark Wells. 2000a. Virtual signing: Capture, animation, storage and transmission-an overview of the visicast project. (2000). (Cited on page 14).

[19] Ursula Bellugi and Susan Fischer. 1972. A comparison of sign language and spoken language. *Cognition* 1, 2 (1972), 173–200. DOI:https://doi.org/10.1016/0010-0277(72)90018-2 (Cited on page 20).

[20] Christopher M Bishop. 2006a. Pattern recognition. *Machine learning* 128, 9 (2006). (Cited on page 6).

[21] Christopher M Bishop. 2006b. *Pattern recognition and machine learning*. springer. (Cited on pages 69, 73).

[22] Andrew Eliot Borthwick. 1999. *A Maximum Entropy Approach to Named Entity Recognition*. Ph.D. Dissertation. USA. Advisor(s) Grishman, Ralph. ISBN:0599472324 AAI9945252. (Cited on page 53).

[23] Mehrez Boulares and Mohamed Jemni. 2019. Automatic hand motion analysis for the sign language space management. *Pattern Analysis and Applications* 22, 2 (2019), 311–341. (Cited on page 12).

[24] Annelies Braffort, Michael Filhol, Maxime Delorme, Laurence Bolot, Annick Choisier, and Cyril Verrecchia. 2016. KAZOO: a sign language generation platform based on production rules. *Universal Access in the Information Society* 15, 4 (2016), 541–550. (Cited on page 14).

[25] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, and others. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 16–31. (Cited on page 14).

[26] Jan Bungeroth, Daniel Stein, Philippe Dreuw, Morteza Zahedi, and Hermann Ney. 2006. A German Sign Language Corpus of the Domain Weather Report.. In *LREC*. 2000–2003. (Cited on page 29).

[27] Tom Calvert. 2016. Approaches to the representation of human movement: notation, animation and motion capture. In *Dance Notations and Robot Motion*. Springer, 49–68. (Cited on page 13).

[28] Carlo Camporesi, Yazhou Huang, and Marcelo Kallmann. 2010. Interactive motion modeling and parameterization by direct demonstration. In *International Conference on Intelligent Virtual Agents*. Springer, 77–90. (Cited on page 18).

[29] Pavel Campr, Marek Hrúz, and Jana Trojanová. 2008. Collection and preprocessing of czech sign language corpus for sign language recognition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*. (Cited on page 30).

[30] Blender Online Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. http://www.blender.org (Cited on page 52).

[31] Stephen Cox, Michael Lincoln, Judy Tryggvason, Melanie Nakisa, Mark Wells, Marcus Tutt, and Sanja Abbott. 2002. Tessa, a System to Aid Communication with Deaf People. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies (ASSETS '02)*. Association for Computing Machinery, New York, NY, USA, 205–212. DOI:10.1145/638249.638287 (Cited on pages 4, 15, 29, 30).

[32] Onno A Crasborn and IEP Zwitserlood. 2008. The Corpus NGT: an online corpus for professionals and laymen. (2008). (Cited on page 29).

[33] David Crystal. 1969. *Prosodic systems and intonation in English*. Vol. 1. CUP Archive. (Cited on page 18).

[34] Anne Cutler, Delphine Dahan, and Wilma Van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and speech* 40, 2 (1997), 141–201. (Cited on pages 18, 19).

[35] Fernando Wagner Da Silva, Luiz Velho, Paulo Roma Cavalcanti, and Jonas Gomes. 1997. An architecture for motion capture based animation. In *Proceedings X Brazilian Symposium on Computer Graphics and Image Processing*. IEEE, 49–56. (Cited on page 13).

[36] Kyle Duarte. 2012. *Motion capture and avatars as portals for analyzing the linguistic structure of signed languages*. Ph.D. Dissertation. Phd thesis, université de bretagne sud. (Cited on pages xvii, 7, 68, 91, 123, 126, 127, 128, 133, 139, 145).

[37] Kyle Duarte and S Gibet. 2011. Presentation of the SignCom project. In *Proceedings of the First International Workshop on Sign Language Translation and Avatar Technology, Berlin, Germany*. 10–11. (Cited on page 29).

[38] Sarah Ebling and John Glauert. 2016a. Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society* 15, 4 (2016), 577–587. DOI:10.1007/s10209-015-0408-1 (Cited on pages 2, 32, 33, 36).

[39] Sarah Ebling and John Glauert. 2016b. Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society* 15, 4 (2016), 577–587. (Cited on page 32).

[40] Ralph Elliott, John RW Glauert, JR Kennaway, and Ian Marshall. 2000. The development of language processing support for the ViSiCAST project. In *Proceedings of the fourth international ACM conference on Assistive technologies*. 101–108. (Cited on page 14).

[41] Ralph Elliott, John RW Glauert, JR Kennaway, Ian Marshall, and Eva Safar. 2008. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society* 6, 4 (2008), 375–391. (Cited on page 14).

[42] Petros Faloutsos, Michiel Van de Panne, and Demetri Terzopoulos. 2001. Composable controllers for physics-based character animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 251–260. (Cited on page 13).

[43] Susan D Fischer, Lorraine A Delhorne, and Charlotte M Reed. 1999. Effects of Rate of Presentation on the Reception of American Sign Language. *Journal of Speech, Language, and Hearing Research* 42, 3 (1999), 568–582. DOI:10.1044/jslhr.4203.568 (Cited on pages 20, 94).

[44] Anthony Fox and others. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford University Press. (Cited on pages 18, 19).

[45] Sylvie Gibet, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour. 2011. The SignCom system for data-driven animation of interactive virtual signers: Methodology and Evaluation. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1, 1 (2011), 1–23. (Cited on page 29).

[46] Michael Gleicher. 2008. More Motion Capture in Games-Can We Make Example-Based Approaches Scale?. In *International Workshop on Motion in Games*. Springer, 82–93. (Cited on page 15).

[47] SignTime GmbH. 2021. SiMAX. (2021). https://simax.media/, last accessed on April 5, 2021. (Cited on pages 95, 120).

[48] Jigar Gohel, Sedeeq Al-khazraji, and Matt Huenerfauth. 2018a. Modeling the Use of Space for Pointing in American Sign Language Animation. (2018). (Cited on pages 12, 13, 17).

[49] Jigar Gohel, Sedeeq Al-khazraji, and Matt Huenerfauth. 2018b. Modeling the Use of Space for Pointing in American Sign Language Animation. (2018). (Cited on page 94).

[50] Angus B Grieve-Smith. 1999. English to American Sign Language machine translation of weather reports. In *Proceedings of the Second High Desert Student Conference in Linguistics (HDSL2), Albuquerque, NM*. 23–30. (Cited on page 28).

[51] François Grosjean. 1977. The perception of rate in spoken and sign languages. *Perception & Psychophysics* 22, 4 (1977), 408–413. (Cited on pages 20, 23).

[52] François Grosjean. 1979. A study of timing in a manual and a spoken language: American Sign Language and English. *Journal of Psycholinguistic Research* 8, 4 (1979), 379–405. DOI:10.1007/BF01067141 (Cited on pages 7, 20, 21, 23, 31, 34, 35, 55, 61, 64, 68, 75).

[53] François Grosjean, Lysiane Grosjean, and Harlan Lane. 1979a. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology* 11, 1 (1979), 58 – 81. DOI:https://doi.org/10.1016/0010-0285(79)90004-5 (Cited on pages 7, 19, 22, 23, 31, 36, 53, 61, 63, 75, 94).

[54] François Grosjean, Lysiane Grosjean, and Harlan Lane. 1979b. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology* 11, 1 (1979), 58 – 81. DOI:https://doi.org/10.1016/0010-0285(79)90004-5 (Cited on pages 97, 112).

[55] François Grosjean and Harlan Lane. 1977a. Pauses and syntax in American sign language. *Cognition* 5, 2 (1977), 101 – 117. DOI:https://doi.org/10.1016/0010-0277(77)90006-3 (Cited on pages 19, 23, 31, 36, 75, 94).

[56] François Grosjean and Harlan Lane. 1977b. Pauses and syntax in American sign language. *Cognition* 5, 2 (1977), 101 – 117. DOI:https://doi.org/10.1016/0010-0277(77)90006-3 (Cited on page 97).

[57] David Halliday, Robert Resnick, and Jearl Walker. 2013. *Fundamentals of physics*. John Wiley & Sons. (Cited on page 17).

[58] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media. (Cited on page 38).

[59] Ian Douglas Horswill. 2009. Lightweight procedural animation with believable physical interactions. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 1 (2009), 39–49. (Cited on page 13).

[60] Matt Huenerfauth. 2008a. Evaluation of a Psycholinguistically Motivated Timing Model for Animations of American Sign Language. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '08)*. Association for Computing Machinery, New York, NY, USA, 129–136. DOI:10.1145/1414471.1414496 (Cited on pages 2, 3, 4, 23, 24, 28, 31, 33, 34, 47, 53, 64, 65, 69, 75, 79, 82, 87, 94, 97, 98, 99).

[61] Matt Huenerfauth. 2008b. Evaluation of a Psycholinguistically Motivated Timing Model for Animations of American Sign Language. In *Proceedings of the 10th International ACM*

SIGACCESS Conference on Computers and Accessibility (ASSETS '08). Association for Computing Machinery, New York, NY, USA, 129–136. DOI:10.1145/1414471.1414496 (Cited on pages 94, 119).

[62] Matt Huenerfauth. 2008c. Spatial, temporal, and semantic models for American Sign Language generation: implications for gesture generation. *International Journal of Semantic Computing* 2, 01 (2008), 21–45. (Cited on page 38).

[63] Matt Huenerfauth. 2009a. A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. *ACM Trans. Access. Comput.* 2, 2, Article 9 (June 2009), 31 pages. DOI:10.1145/1530064.1530067 (Cited on pages 2, 4, 7, 13, 20, 22, 23, 24, 28, 31, 33, 34, 35, 36, 47, 53, 64, 65, 75, 76, 79, 82, 91, 93, 94, 97, 98, 99, 121, 123, 139, 145).

[64] Matt Huenerfauth. 2009b. A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. *ACM Trans. Access. Comput.* 2, 2, Article 9 (June 2009), 31 pages. DOI:10.1145/1530064.1530067 (Cited on pages 94, 95, 119, 120).

[65] Matt Huenerfauth. 2014. Learning to generate understandable animations of American Sign Language. (Cited on pages 2, 28, 31).

[66] Matt Huenerfauth and Vicki Hanson. 2009. Sign language in the interface: access for deaf signers. *Universal Access Handbook. NJ: Erlbaum* 38 (2009). (Cited on page 14).

[67] Matt Huenerfauth and Hernisa Kacorri. 2014a. Release of experimental stimuli and questions for evaluating facial expressions in animations of American Sign Language. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, The 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland*. (Cited on page 38).

[68] Matt Huenerfauth and Hernisa Kacorri. 2014b. Release of Experimental Stimuli and Questions for Evaluating Facial Expressions in Animations of American Sign Language. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, The 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland.* (Cited on pages 119, 120).

[69] Matt Huenerfauth and Hernisa Kacorri. 2015a. Augmenting EMBR virtual human animation system with MPEG-4 controls for producing ASL facial expressions. In *International Symposium*

on Sign Language Translation and Avatar Technology, Vol. 3. (Cited on pages 7, 20, 91, 93, 121, 123, 139, 145).

[70] Matt Huenerfauth and Hernisa Kacorri. 2015b. Augmenting EMBR virtual human animation system with MPEG-4 controls for producing ASL facial expressions. In *International Symposium on Sign Language Translation and Avatar Technology*, Vol. 3. (Cited on page 94).

[71] Matt Huenerfauth and Pengfei Lu. 2010. Modeling and synthesizing spatially inflected verbs for American sign language animations. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10*. ACM Press. DOI:10.1145/1878803.1878823 (Cited on page 119).

[72] Matt Huenerfauth and Pengfei Lu. 2011. Effect of spatial reference and verb inflection on the usability of sign language animations. *Universal Access in the Information Society* 11, 2 (Sept. 2011), 169–184. DOI:10.1007/s10209-011-0247-7 (Cited on pages 119, 120).

[73] Matt Huenerfauth and Pengfei Lu. 2012. Effect of spatial reference and verb inflection on the usability of sign language animations. *Universal Access in the Information Society* 11, 2 (2012), 169–184. DOI:10.1007/s10209-011-0247-7 (Cited on page 3).

[74] Matt Huenerfauth, Liming Zhao, Erdan Gu, and Jan Allbeck. 2007. Evaluating American Sign Language Generation through the Participation of Native ASL Signers. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '07)*. Association for Computing Machinery, New York, NY, USA, 211–218. DOI:10.1145/1296843.1296879 (Cited on page 81).

[75] Matt Huenerfauth, Liming Zhao, Erdan Gu, and Jan Allbeck. 2008. Evaluation of American sign language generation by native ASL signers. *ACM Transactions on Accessible Computing (TACCESS)* 1, 1 (2008), 1–27. (Cited on page 38).

[76] Spandana Jaggumantri, Sedeeq Al-khazraji, Abraham Glasser, and Matt Huenerfauth. 2019. Designing an Interface to Support the Creation of Animations of Individual ASL Signs. (2019). (Cited on page 17).

[77] Vince Jennings, Ralph Elliott, Richard Kennaway, and John Glauert. 2010. Requirements for a signing avatar. In *Proc. Workshop on Corpora and Sign Language Technologies (CSLT), LREC*. 33–136. (Cited on pages 32, 33, 34, 35).

[78] JMP. 1989-2019. SAS Institute Inc. Cary, NC. (1989-2019). (Cited on page 70).

[79] Robert E Johnson and Scott K Liddell. 2011. A segmental framework for representing signs phonetically. *Sign Language Studies* 11, 3 (2011), 408–463. (Cited on pages 125, 126).

[80] Ollie Johnston and Frank Thomas. 1981a. *The illusion of life: Disney animation*. Disney Editions New York. (Cited on page 18).

[81] Ollie Johnston and Frank Thomas. 1981b. *The illusion of life: Disney animation*. Disney Editions New York. (Cited on pages 95, 119).

[82] Hernisa Kacorri. 2016. Data-Driven Synthesis and Evaluation of Syntactic Facial Expressions in American Sign Language Animation. (2016). (Cited on pages 12, 15, 86).

[83] Hernisa Kacorri and Matt Huenerfauth. 2016. Continuous profile models in ASL syntactic facial expression synthesis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2084–2093. (Cited on page 3).

[84] Hernisa Kacorri, Matt Huenerfauth, Sarah Ebling, Kasmira Patel, Kellie Menzies, and Mackenzie Willard. 2017. Regression Analysis of Demographic and Technology-Experience Factors Influencing Acceptance of Sign Language Animation. *ACM Trans. Access. Comput.* 10, 1, Article 3 (April 2017), 33 pages. DOI:10.1145/3046787 (Cited on pages xiv, 77, 86).

[85] Hernisa Kacorri, Pengfei Lu, and Matt Huenerfauth. 2013a. Effect of displaying human videos during an evaluation study of American Sign Language animation. *ACM Transactions on Accessible Computing (TACCESS)* 5, 2 (2013), 1–31. (Cited on page 38).

[86] Hernisa Kacorri, Pengfei Lu, and Matt Huenerfauth. 2013b. Effect of displaying human videos during an evaluation study of American Sign Language animation. *ACM Transactions on Accessible Computing (TACCESS)* 5, 2 (2013), 1–31. (Cited on page 99).

[87] Hernisa Kacorri, Pengfei Lu, and Matt Huenerfauth. 2013c. Evaluating facial expressions in American Sign Language animations for accessible online information. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 510–519. (Cited on pages 38, 86).

[88] Avi Kak. 2002. Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language. IEEE Computer Society, 167–172. ISBN:9780769518343;0769518346; (Cited on page 30).

[89] Abhishek Kannekanti, Sedeeq Al-khazraji, and Matt Huenerfauth. 2019. Design and Evaluation of a User-Interface for Authoring Sentences of American Sign Language Animation. In *Universal Access in Human-Computer Interaction. Theory, Methods and Tools*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 258–267. ISBN:978-3-030-23560-4 (Cited on page 17).

[90] Sarah Kelsall. 2006. Movement and location notation for American sign language. (2006). (Cited on page 12).

[91] Duane Knudson. 2007. *Fundamentals of biomechanics*. Springer Science & Business Media. (Cited on page 18).

[92] Ron Kohavi and others. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, Vol. 14. Montreal, Canada, 1137–1145. (Cited on page 38).

[93] Motion Light Lab. 2021. Dimensions - Meet Zoe. (2021). https://www.motionlightlab.com/dimensions, last accessed on April 5, 2021. (Cited on pages 95, 120).

[94] John Lasseter. 1987. Principles of traditional animation applied to 3D computer animation. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. 35–44. (Cited on page 18).

[95] Scott K Liddell and Robert E Johnson. 1989. American sign language: The phonological base. *Sign language studies* 64, 1 (1989), 195–277. (Cited on pages xiii, 12, 13, 129).

[96] Zicheng Liu and Michael F Cohen. 1995. Keyframe motion optimization by relaxing speed and timing. In *Computer Animation and Simulation'95*. Springer, 144–153. (Cited on page 18).

[97] Pengfei Lu. 2014. *Data-driven synthesis of animations of spatially inflected American Sign Language verbs using human data*. Ph.D. Dissertation. (Cited on page 3).

[98] Pengfei Lu and Matt Huenerfauth. 2010. Collecting a motion-capture corpus of American Sign Language for data-driven generation research. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*. Association for Computational Linguistics, 89–97. (Cited on page 41).

[99] Pengfei Lu and Matt Huenerfauth. 2012. Cuny american sign language motion-capture corpus: first release. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, The 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey*. (Cited on pages 41, 49).

[100] Pengfei Lu and Matt Huenerfauth. 2014a. Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation. *Computer Speech Language* 28, 3 (2014), 812 – 831. DOI:https://doi.org/10.1016/j.csl.2013.10.004 (Cited on pages 3, 30, 41, 49, 76, 106).

[101] Pengfei Lu and Matt Huenerfauth. 2014b. Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation. *Computer Speech & Language* 28, 3 (2014), 812 – 831. DOI:https://doi.org/10.1016/j.csl.2013.10.004 (Cited on pages 111, 112).

[102] Pengfei Lu and Hernisa Kacorri. 2012. Effect of presenting video as a baseline during an American Sign Language animation user study. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 183–190. (Cited on page 99).

[103] John McDonald, Rosalee Wolfe, Jerry Schnepp, Julie Hochgesang, Diana Gorman Jamrozik, Marie Stumbo, Larwan Berke, Melissa Bialek, and Farah Thomas. 2016. An automated technique for real-time production of lifelike animations of American Sign Language. *Universal Access in the Information Society* 15, 4 (2016), 551–566. (Cited on page 94).

[104] Ross E. Mitchell. 2005. How Many Deaf People Are There in the United States? Estimates From the Survey of Income and Program Participation. *The Journal of Deaf Studies and Deaf Education* 11, 1 (09 2005), 112–119. (Cited on page 1).

[105] Mark Mizuguchi, John Buchanan, and Tom Calvert. 2001. Data driven motion transitions for interactive games. In *Eurographics 2001 Short Presentations*, Vol. 2. 6. (Cited on page 15).

[106] Sara Morrissey and Andy Way. 2005. An Example-Based Approach to Translating Sign Language. In *MT Summit X*. Citeseer, 109. (Cited on pages 29, 45).

[107] Luciana Porcher Nedel and Daniel Thalmann. 1998a. Modeling and deformation of the human body using an anatomically-based approach. In *Proceedings Computer Animation'98 (Cat. No. 98EX169)*. IEEE, 34–40. (Cited on page 13).

[108] Luciana Porcher Nedel and Daniel Thalmann. 1998b. Modeling and deformation of the human body using an anatomically-based approach. In *Proceedings Computer Animation'98 (Cat. No. 98EX169)*. IEEE, 34–40. (Cited on page 13).

[109] Carol Neidle. 2001. SignStream™: A database tool for research on visual-gestural language. *Sign language & linguistics* 4, 1-2 (2001), 203–214. (Cited on page 41).

[110] Carol Jan Neidle, Judy Kegl, Benjamin Bahan, Dawn MacLaughlin, and Robert G Lee. 2000. *The syntax of American Sign Language: Functional categories and hierarchical structure*. MIT press. (Cited on page 30).

[111] Sega of America. 2021. Sonic The Hedgehog Website. (2021). https://www.sonicthehedgehog.com/, last accessed on April 11, 2021. (Cited on pages 95, 119).

[112] World Federation of the Deaf. Our Work. 2020. Our Work. (2020). http://wfdeaf.org/our-work/, last accessed on May 15, 2020. (Cited on page 1).

[113] Alok Parlikar and Alan W Black. 2012a. Data-driven phrasing for speech synthesis in low-resource languages. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4013–4016. (Cited on page 27).

[114] Alok Parlikar and Alan W Black. 2012b. Modeling pause-duration for style-specific speech synthesis. In *Thirteenth Annual Conference of the International Speech Communication Association*. (Cited on page 27).

[115] Afra Pascual, Mireia Ribera, and Toni Granollers. 2014. Impact of Web Accessibility Barriers on Users with Hearing Impairment. In *Proceedings of the XV International Conference on Human Computer Interaction*. Association for Computing Machinery, New York, NY, USA, Article 8, 2 pages. DOI:10.1145/2662253.2662261 (Cited on page 2).

[116] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830. (Cited on page 69).

[117] David M Perlmutter. 1993. Sonority and syllable structure in American Sign Language. In *Current issues in ASL phonology*. Elsevier, 227–261. (Cited on page 12).

[118] Roland Pfau, Markus Steinbach, and Bencie Woll. 2012. *Sign language: An international handbook*. Vol. 37. Walter de Gruyter. (Cited on pages 7, 31, 36, 61, 63).

[119] Simorin Pinto. 2021. New virtual avatar "Star" to bring books to life through sign language. (2021). https://channels.theinnovationenterprise.com/articles/new-virtual-avatar-star-to-bring-books-to-life-through-sign-language, last accessed on April 5, 2021. (Cited on pages 95, 120).

[120] Wendy Sandler. 2010. Prosody and syntax in sign languages. *Transactions of the philological society* 108, 3 (2010), 298–328. (Cited on pages 18, 19).

[121] Wendy Sandler. 2011. *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language*. Vol. 32. Walter de Gruyter. (Cited on page 12).

[122] Wendy Sandler. 2012. The phonological organization of sign languages. *Language and linguistics compass* 6, 3 (2012), 162–182. (Cited on pages 18, 19).

[123] Alexander Savenko. 2002. *Animating character locomotion using biomechanics based figure models*. Ph.D. Dissertation. De Montfort University. (Cited on page 13).

[124] scikit learn. 2020. sklearn.ensemble.GradientBoostingRegressor. (2020). https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html, last accessed on MAY 12, 2020. (Cited on page 73).

[125] SEGA. 2021. Sonic The Hedgehog™ | SEGA. (2021). https://www.sega.com/games/sonic-hedgehog, last accessed on April 11, 2021. (Cited on pages 95, 119).

[126] Jérémie Segouat and Annelies Braffort. 2009. Toward the study of sign language coarticulation: methodology proposal. In *2009 Second International Conferences on Advances in Computer-Human Interactions*. IEEE, 369–374. (Cited on pages 2, 4, 32, 33, 35).

[127] J. Segouat and A. Braffort. 2009. Toward the Study of Sign Language Coarticulation: Methodology Proposal. In *2009 Second International Conferences on Advances in Computer-Human Interactions*. 369–374. (Cited on page 15).

[128] Ari Shapiro. 2011. Building a character animation system. In *International conference on motion in games*. Springer, 98–109. (Cited on page 18).

[129] Ari Shapiro, Derek Chu, Brian Allen, and Petros Faloutsos. 2007. A dynamic controller toolkit. In *Proceedings of the 2007 ACM SIGGRAPH symposium on Video games*. 15–20. (Cited on page 13).

[130] ELAN (Version 5.2) [Computer software]. 2020. Nijmegen: Max Planck Institute for Psycholinguistics. (2020). https://tla.mpi.nl/tools/tla-tools/elan/, last accessed on May 13, 2020. (Cited on pages 5, 40, 45, 58, 143).

[131] Arda Söylev and Engin Mendi. 2014. Turkish Sign Language Animation with motion capture. In *2014 22nd Signal Processing and Communications Applications Conference (SIU)*. IEEE, 834–837. (Cited on page 4).

[132] D Stein, J Bungeroth, and H Ney. 2006. Morpho-syntax based statistical methods for sign language translation. In *11th Annual conference of the European Association for Machine Translation, Oslo, Norway*. Citeseer, 169–177. (Cited on page 28).

[133] Sign Smith Studio. 2020. Vcom3D. (2020). http://www.vcom3d.com/signsmith.php, last accessed on May 15, 2020. (Cited on pages 13, 15, 18, 27, 32, 33, 34, 36, 82, 99, 102, 134).

[134] Sign Smith Studio. 2021. Vcom3D. (2021). http://www.vcom3d.com/, last accessed on April 3, 2021. (Cited on page 111).

[135] Kara Technologies. 2021. Kara Technologies Homepage. (2021). https://www.kara.tech, last accessed on April 11, 2021. (Cited on page 119).

[136] Jorge Toro. 2004. Automated 3D animation system to inflect agreement verbs. In *Proc. 6th High Desert Linguistics Conf*. (Cited on pages 30, 45).

[137] Carol Bloomquist Traxler. 2000. The Stanford Achievement Test, 9th Edition: National Norming and Performance Standards for Deaf and Hard-of-Hearing Students. *The Journal of Deaf Studies and Deaf Education* 5, 4 (09 2000), 337–348. (Cited on page 2).

[138] H. Van Welbergen, B. J. H. Van Basten, A. Egges, Zs. M. Ruttkay, and M. H. Overmars. 2010. Real Time Animation of Virtual Humans: A Trade-off Between Naturalness and Control. *Computer Graphics Forum* 29, 8 (2010), 2530–2554. DOI:10.1111/j.1467-8659.2010.01822.x (Cited on page 15).

[139] Christian Vogler and Carol Neidle. 2012. A new web interface to facilitate access to corpora: development of the ASLLRP data access interface. (2012). https://open.bu.edu/handle/2144/31886 (Cited on page 20).

[140] Ann Wennerstrom. 2001. *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press. (Cited on page 19).

[141] Harold Whitaker and John Halas. 2013. *Timing for animation*. CRC Press. (Cited on page 18).

[142] Wikipedia. 2020. Cross-validation (statistics). (2020). https://en.wikipedia.org/wiki/Cross-validation_(statistics), last accessed on MAY 15, 2020. (Cited on page 38).

[143] Ronnie B. Wilbur. 2009. Effects of Varying Rate of Signing on ASL Manual Signs and Nonmanual Markers. *Language and Speech* 52, 2-3 (2009), 245–285. DOI:10.1177/0023830909103174 (Cited on pages 7, 21, 23, 31, 35, 61, 64).

[144] Wayne L Wooten and Jessica K Hodgins. 2000. Simulating leaping, tumbling, landing and balancing humans. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, Vol. 1. IEEE, 656–662. (Cited on page 13).

[145] Pawel Wrotek, Odest Chadwicke Jenkins, and Morgan McGuire. 2006. Dynamo: dynamic, data-driven character control with adjustable balance. In *Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames*. 61–70. (Cited on page 15).

[146] VM Zatsiorsky. 1998. Kinematics of human motion, Human Kinetics. *Urbana Champaign* (1998). (Cited on page 124).

[147] Yi Zhang, Susan Finger, and Stephannie Behrens. 2003. *Introduction to mechanisms*. Carnegie Mellon University. (Cited on page 17).

Appendices

Appendix A

Appendix for Interview Study

This is an appendix for Chapter 6: "Model Evaluation."

A.1 Simple of the Selected Stories for the 2008 Model and ASL-Speed Comparison

- **Signer SIA02 - Story44:** MOVIE IX-1-s:S WATCH MOVIE RECENTLY IX-1-s:1 OLD MOVIE BUT IX-1-s:S SAW FIRST TIME NAME REALLY COOL FRIEND SEE MOVIE BIG IX-1-s:S cl"DON'T SAY #TOO LONG IX-1-s:S TWO HOURS IX-1-s:S NOT HAVE TIME WATCH THAT TWO HOURS THREE HOURS IX-1-s:S DO-DO+ BUSY THREE HOUR WHY LONG FOR IX-1-s:S TO SIT ONE PLACE MOVIE WATCH 3-D IX-1-s:S NOW IX-1-s:S WISH IX-1-s:S #BACK WISH WATCH 3-D THAT MOVIE THAT MOVIE REALLY NICE REALLY IX-1-s:1 COMPARE AND fs-U AMERICA IX-1-s:3 HUMAN POSS-s:3 IX-1-s:3 JOIN POSS-s:2 WORLD IX-1-s:3 3:STEAL:2 POSS-s:2 BODY 3:STEAL:2 POSS-s:2 CLOTHES REALLY TALL HUMAN CL"SMALL" CL"CHARACTER/ACTION IX-1-s:2 ZOOM JUMP LIVE IN TREE THINK SIMILAR ALMOST SAME IX-1-s:4 AMERICA NATIVE AMERICA NATIVE HERE LIVE HERE AMERICA BRITISH AMERICA TAKE-OVER EXPELL POSS-s:4 COUNTRY REALLY SAME SIMILAR REALLY NICE SPECIAL fs-EFFECT CL"Flying/jumping STORY GOOD MAKE IX-1-s:S CRY SENSITIVE CONNECT FALL-LOVE LOVE HAVE WAR HAVE FIGHT HAVE EVERYTHING VARIETY EMOTION

IN THAT FUNNY LIST-TWO EXCITING WAR LIST-ALL CHALLENGE PLUS NATURE ALMOST SAME fs:ABYSS WHERE CONNECTION WITH EARTH CONNECT WITH ENERGRY THAT NICE IX-1-s:S ENJOY THAT MOVIE IX-1-s:S WATCH SECOND TIME WILL IX-1-s:S DON'T-MIND

- **Signer SIA02 -Story43:** #HS VERSUS COLLEGE COLLEGE INDEPENDENT ON POSS-s:A #OWN #HS úmm' HAVE ? HAVE RULE COLLEGE ON #OWN IX-1-s:A 'umm" #DO WANT THINKŜELF:A #WHAT IX-1-s:A RISK UP-TO-YOU:A #HS IX-1-s:A LESS IX-1-s:A HAVE CONCLUSION HAVE LIST-ALL COLLEGE ON #OWN umm COLLEGE 0:PAY:2 #HS NOT BUT SOME 0:PAY:1 SAME WORK++ 0:LOOK:2 COLLEGE DEGREE #HS DEGREE 0:INGORE:1 0:LOOK:2 COLLEGE ONLY IX-1-s:S PREFER COLLEGE OF-COURSE IX-1-s:S RECENT GRADUATE COLLEGE PREVIOUS #DEC HAVE fs-PE DEGREE

- **Signer SIB01 - Story17:** OH FINE POSS-s:S TWO PREFER fs-PHONE FIRST POSS-s:S EXPERIENCE WHAT LIST-ONE fs-BLACKBERRY IX-1-s:S HAVE #IT FOR MANY+ YEAR IX-1-s:S START BLACK fs-BERRY WHEN IX-1-s:S ENTER MIDDLE SCHOOL IX-1-s:S #OR HIGH-SCHOOL IX-1-s:S START PROCESS USE #THEN IX-1-s:S REALIZE fs-MAC HAVE INVOLVE fs-IPHONE INVOLVE LIST-TWO CAMERA LIST-THREE GAME LIST-TWO INTERNET THROUGH fs-SURF LIST-TWO A-LOT #OF ACCESS THAN BLACKBERRY SO IX-1-s:S TRANSFER IX-1-s:1 fs-IPHONE AND IX-1-s:S LOVE #IT HAVE ONE-HUNDRED PERCENT ACCESS FOR IX-1-s:S #EMAIL LIST-TWO #TEXT LIST-THREE INTERNET LIST-TWO fs-VIDEO LIST-TWO ETC PLEASE DON'T GO-OUT FOR BLACKBERRY GO-OUT FOR fs-IPHONE

- **Signer SIB01 - Story27:** HOPE ARTICLE HOPE fs-DIAMOND THAT LONG-TIME-AGO UMM ONE FAMILY WHO COLLECT fs-GEMS FOR MANY YEARS HAPPEN PASS-ALONG FOR MANY YEARS TRUE-BUSINESS ONE PERSON THINK RESEARCH IX-1-s:1 #WAS STEAL THAT DIAMOND CL"ROCK" LONG-AGO FAMILY #WAS POOR FROM QUOTE FRANCE fs:JEWEL CL"DEBATE/PROBELMS" #SO THAT DIAMOND WAS NEVER GIVEN TO OTHER PEOPLE PUT TO MUSEUM WITH CL"THICK GLASS CL"DESCRIBING HERE IX-1-s:S GUESS PEOPLE LOOK WATCH THAT DIAMOND THAT DIAMOND WORTH FORTUNE SO IX-1-s:S THINK PROBABLY SOLD TO MUSEUM THAT WORTH TWO MILLION DOLLAR NOT SURE fs-BIG DIAMOND PRETTY DIAMOND IX-1-s:S WANT #IT IX-1-s:S

- **Signer SIC02 - Story06:** BUT MY FAVORITE TEAM fs:LAKERS OF-COURSE CHAMPION++++ MANY CHAMPION++ SET-ASIDE BUT fs:DRAFT IX-1-s:1 PICK LATE LAST BOTTOM BUT ALWAYS GOOD TRADE PICK+++ EXCELLENT ONE PLAYER TO THINK UM EXCUSE WHY IX-1-s:2 PUSH TO fs-KOBE CL"BALL HAND-BROKEN FOREVER CL"broken WILL CL"HAND NAMED SERIOUS CL"PAIN FOREVER NEED HURT REST BUT PLAY+++ CHAMPION WILL CL"HAND USE CL"PROTECTION NOW-ON READ RECENT READ ARTICLE AND fs-DRAFT BUT TRADING TODAY fs:JULY FIRST KNOW WHO FAMOUS BECOME FREE fs-AGENT TO-DAY CL"COMMOTION" NEED TRANSFER TO fs-LAKERS SHOULD BUT fs:LAKERS CAN'T AFFORD IX-1-s:3 MONEY LIMIT MILLION EACH-YEAR MAXIMUM fs-LAKERS LIMIT fs-CAP FINISH COLLECT HIGH fs-KNICKS CAN AFFORD LIST-THREE CHICAGO LIST-ONE CAN AFFORD MAYBE MIAMI CAN AFFORD BUT IX-1-s:3 BEST STAY WITH CLEVELAND BUT COACH NOW COACH FIRED NEW COACH HIRE NEW COACH IX-1-s:4 QUIT DON'T-WANT CL"RUMORS/DEBATES" #SO MAYBE fs:KO fs:LBJ WANT LEAVE IX-1-s:3 NOW TALK ABOUT WHEN DEADLINE THAT GOOD QUESTION DEADLINE THAT POINT GET-CLOSER LAST THAT IMPORTANT NEED DECIDE NOW OPEN FOR TALK REALLY NOTHING NONE DIFFER-ENT

- **Signer SIC02 - Story21:** MOVIE IX-1-s:S WATCH MOVIE RECENTLY IX-1-s:1 OLD MOVIE BUT IX-1-s:S SAW FIRST TIME NAME REALLY COOL FRIEND SEE MOVIE BIG IX-1-s:S cl"DON'T SAY #TOO LONG IX-1-s:S TWO HOURS IX-1-s:S NOT HAVE TIME WATCH THAT TWO HOURS THREE HOURS IX-1-s:S DO-DO+ BUSY THREE HOUR WHY LONG FOR IX-1-s:S TO SIT ONE PLACE MOVIE WATCH 3-D IX-1-s:S NOW IX-1-s:S WISH IX-1-s:S #BACK WISH WATCH 3-D THAT MOVIE THAT MOVIE REALLY NICE REALLY IX-1-s:1 COMPARE AND fs-U AMERICA IX-1-s:3 HUMAN POSS-s:3 IX-1-s:3 JOIN POSS-s:2 WORLD IX-1-s:3 3:STEAL:2 POSS-s:2 BODY 3:STEAL:2 POSS-s:2 CLOTHES REALLY TALL HUMAN CL"SMALL" CL"CHARACTER/ACTION IX-1-s:2 ZOOM JUMP LIVE IN TREE THINK SIMILAR ALMOST SAME IX-1-s:4 AMERICA NATIVE AMERICA NATIVE HERE LIVE HERE AMERICA BRITISH AMERICA TAKE-OVER EXPELL POSS-s:4 COUNTRY REALLY SAME SIMILAR REALLY NICE SPECIAL fs-EFFECT CL"Flying/jumping STORY GOOD MAKE IX-1-s:S CRY SENSITIVE CON-NECT FALL-LOVE LOVE HAVE WAR HAVE FIGHT HAVE EVERYTHING VARIETY EMOTION IN THAT FUNNY LIST-TWO EXCITING WAR LIST-ALL CHALLENGE PLUS NATURE ALMOST SAME fs:ABYSS WHERE CONNECTION WITH EARTH CONNECT WITH ENERGRY THAT

NICE IX-1-s:S ENJOY THAT MOVIE IX-1-s:S WATCH SECOND TIME WILL IX-1-s:S DON'T-MIND

A.2 ASLSPEED2018 Survey Recruitment Flyer

R·I·T

Advertisement for Research Study
Earn $40 for a one-hour appointment
Open to people who used ASL since childhood

When you were a child, did you use American Sign Language at home?

When you were a child (before age 7), did you use American Sign Language at school?

If you answer "yes" to one of these questions, you are invited to participate in a research study at Rochester Institute of Technology. Participants will view animations or videos of sign language for one hour on a computer, and they will be paid $40 for their participation.

The computer animations are of American Sign Language sentences, and participants will fill out a survey about whether the animations are correct and understandable.

To make an appointment, please contact:

Mr. Peter Yeung, Research Assistant
Linguistic and Assistive Technologies Laboratory
Rochester Institute of Technology
Email: latlabrit**2**@gmail.com

For general information about the project, please contact:

Dr. Matt Huenerfauth, Professor
Department of Information Sciences and Technologies
Rochester Institute of Technology (RIT)
Email: matt.huenerfauth@rit.edu

A.3 ASLSPEED2018 Study Information Handout

Project Title: Generating Accurate, Understandable Sign Language Animations Based on Analysis of Human Signing

Investigator: Matt Huenerfauth, Associate Professor, Department of Information Sciences and Technologies, Rochester Institute of Technology

You are being asked to participate in a research project. This information is provided so that you can decide whether to participate.

Nature and Purpose of the Project: The purpose of this research project is to evaluate the quality of a computer system that creates American Sign Language (ASL) animations.

Explanation of Procedures: Today, you will be asked to look at a computer screen that will display animations of a 3D human character performing ASL or a video of a human performing ASL. You will be asked to evaluate the understandability and grammatical-correctness of the animation. Any computer devices used will be demonstrated to you ahead of time. Your time is not expected to exceed 60 minutes.

Potential Discomfort and Risks: The potential risks in this project are minimal. The various computer devices are not capable of causing physical harm.

Potential Benefits: You will not receive any direct benefit from participating in this study.

Costs/Reimbursements: You will receive $40 compensation for being in this study.

Alternatives to Participation: You may decide not to participate if you wish.

Termination of Participation: If for some reason you are unable to complete the survey or if there is a technical problem with the computer during your session, the investigator may terminate your participation in the study.

Confidentiality: Every attempt will be made by the investigators to maintain all information collected in this study strictly confidential, except as may be required by court order or law. Authorized representatives of Rochester Institute of Technology, including members of the Institutional Review Board (IRB), a committee charged with protecting the rights and welfare of research subjects, may be provided access to research records.

Withdrawal from the Project: Your participation in this research project is completely voluntary. You may decide to stop participating in this project at any time without penalty. You are free to leave at any time. Refusal to participate in this study or withdrawal from this study will have no effect on any services you may otherwise be entitled to.

Whom to Contact with Questions:
This project has been reviewed by the RIT Institutional Review Board. If you have any questions about your rights as a research participant, you may contact:

Heather Foti, Associate Director, Office of Human Subjects Research
Phone Number: 585-475-7673, Email: hmfsrs@rit.edu

If you have concerns or questions about the conduct of this research project you may contact:

Dr. Matt Huenerfauth, Associate Professor
Department of Information Sciences and Technologies, Rochester Institute of Technology
Departmental Phone Number: 585-475-7924

A.4 ASLSPEED2018 Interview Demographic Paper for Participants

Code: _____
Date: _____
ASLSPEED-2018-Interview

Name: _____

Email or pager: _____

1. What is your gender? (please circle): **Male, Female, other: _____**
2. How old are you? _____
3. Which describes you best? **hearing, hard-of-hearing, deaf/Deaf, other: _____**
4. When did you become deaf/Deaf or hard-of-hearing? _____
5. When did you first learn ASL? (How old were you?) _____
6. Did your parents use ASL at home? **Yes No**
7. Are your parents deaf? **Yes No**
8. Other deaf family? Please explain: _____
9. Your childhood school: (You can circle more than one.)
   residential school for Deaf students,
   daytime school for Deaf students,
   mainstream school
10. In your childhood school, did you use ASL? **Yes No**
    If you used ASL at school, how old were you? _____
11. Education: Did you graduate high school? **Yes No**
    Did you graduate college? **Yes No**
    Did you get a bachelor's degree? **Yes No**
    Did you get a graduate degree? **Yes No**
12. Did you ever use ASL at a college or a university? **Yes No**
13. What language do you use at home? **English, ASL, other: _____**
    (You can circle more than one.)
14. What language do you use at work/school? **English, ASL, other: _____**
    (You can circle more than one.)
15. Please list other connections you have to the deaf/Deaf community. For example: husband, wife, sweetheart, friends, sports, clubs…

    _____

    _____

	How often do you do this?	Never	Monthly	Weekly	Once a day	Several times a day	Always
16.	Watch **TV shows, movies, etc**., on computer, laptop, tablet, or smartphone						
17.	Watch **video clips** on the computer, laptop, tablet, or smartphone						
18.	Download **media files from other people** on the computer, laptop, tablet, or smartphone						
19.	Share **your own media files** on a computer, laptop, tablet, smartphone						

Now, the researcher will show you some animations and have a conversation with you, while taking notes.

After you are finished watching animations and having a conversation with the researcher, please answer these final questions….

	Do you agree or disagree?	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
20.	Computer animations of sign language could be used to give information on a website.					
21.	Computer animations of sign language could be used to give information in a public place (e.g. airport, train station).					
22.	Computer animations of sign language could be used as an interpreter in a face-to-face meeting.					
23.	Computer animations of sign language could be used as an interpreter for telephone relay.					
24.	I would enjoy using computer animations of sign language.					
25.	Other people would enjoy using computer animations of sign language.					

What did you like about the computer animations? What should be improved?

_____

_____

_____

_____

Thank you!

A.5 ASLSPEED2018 Interview Plan and Questions

INTERVIEW CHECKLIST (Take notes during the interview.)

1. INFORMED CONSENT HANDOUT

2. FIRST TWO PAGES OF DEMOGRAPHIC PAPER

3. You begin the interview with some warm-up questions, about whether they had seen animations of ASL before, and what they had thought of them.

4. Begin the interview by showing them the first pair of animations (new2018_... "with pauses and speed changes" and baseline_..."without pauses and speed changes"). Do not tell the participant what is different. Just ask them to watch them, then you can begin by asking some general questions about what they noticed. If they don't seem to notice the speed/pauses changes, then you can begin asking them some subtle questions to guide them to this, e.g. you can draw their attention to a specific few words, mention something about speed, something about pauses, etc. This should be done progressively, and after they note the difference, ask them what they think.

Which story did you show them first here? 1 4 9
Note: You can rotate the order in which you show people the three stories during your interview sessions. Some people can see 1-4-9, other people can see 4-9-1 or 1-9-4.

What difference did you notice between the two videos you just saw? How exactly?

5. Deeper discussion about the animation – focus on the **new2018_ animation here**. Ask them to take a careful look at it.

Ask them to give you feedback about how we could improve it.

Ask if there are particular words they notice.

Is it easy to tell where the sentences begin and end?

Ask them to give you a clear answer about which version they prefer, if they had to choose between the two animations.

Is it too fast? Too slow?

What part of the animation do you think is similar to real human?

How natural are the speed of the animations in the video? Why? Suggestion for improvement?

How natural are the pausing of the animations? Why? Suggestion for improvement?

Which one of the following video you like? Why?

6. Broaden the discussion to focus on **animations of ASL in general**, not just this specific one.

What types of speed or pauses would they like to see in animations like this in general. Are you worried about them being too fast or too slow? If other animations looked like this, would it be good?

Are there times when they would want more robotic signing or more natural signing?

Do they think pauses and speed are important?

If they try to sign these sentences themselves, how is it different?

What do you like about computer animations, in general?

What do you think about ASL animations?

What do you think is the most important characteristics for computer animation of ASL? Why?

(As needed, you can explain that our lab is NOT trying to use these to replace interpreters. The idea is that companies usually have websites in lots of languages, but never ASL because videos of real people are too hard to keep updated. We're trying to make it easier and cheaper for companies to put ASL on websites, but a Deaf person or an interpreter would still need to write the message that the animation shows. It would just be easier for the animation message to be updated, rather than videos of people. Our goal is to get websites of companies and governments around the world to have an ASL version, just like they have English and Spanish versions.)

7. Broaden the discussion, to focus on speed and pauses in **HUMAN** signing.

How does the participant use speed or pauses themselves?

What speed or pausing do they look for in the signing of other people.

If they see someone who is pausing in weird places or has unusual speed changes, do they seem like a "beginner" signer? Or a child? What would they thing if this stuff was wrong.

8. Now, you can show them the next pair of animations (for another story). You can repeat some of the discussions in item 5 above.

Which story pair did you show them second? 1 4 9

While they may look at both the baseline_ animation (without pauses and speed changes) and the new2018_ animation (with pauses and speed changes), for your discussion here, you should ask them to focus on the **new2018_ animation here**.

Ask them to take a careful look at it.

Ask them to give you feedback about how we could improve it.

Ask if there are particular words they notice.

Is it easy to tell where the sentences begin and end?

Ask them to give you a clear answer about which version they prefer, if they had to choose between the two animations.

Is it too fast? Too slow?

What part of the animation do you think is similar to real human?

How natural are the speed of the animations in the video? Why? Suggestion for improvement?

How natural are the pausing of the animations? Why? Suggestion for improvement?

Which one of the following video you like? Why?

9. Then, you can show them the third pair of animations (for the third story. You can repeat some of the discussions in item 5 above.

Which story pair did you show them third? 1 4 9

While they may look at both the baseline_ animation (without pauses and speed changes) and the new2018_ animation (with pauses and speed changes), for your discussion here, you should ask them to focus on the **new2018_ animation here**.

Ask them to take a careful look at it.

Ask them to give you feedback about how we could improve it.

Ask if there are particular words they notice.

Is it easy to tell where the sentences begin and end?

Ask them to give you a clear answer about which version they prefer, if they had to choose between the two animations.

Is it too fast? Too slow?

What part of the animation do you think is similar to real human?

How natural are the speed of the animations in the video? Why? Suggestion for improvement?

How natural are the pausing of the animations? Why? Suggestion for improvement?

Which one of the following video you like? Why?

10. Finally, you can **wind-down the interview**. You can ask about any other feedback they have. You can ask them to mention something they likes and something we should update?

11. And then get them to sign a receipt when you pay them $40 at the end.

Appendix B

Appendix for ASL-Speed 2020 Study

This is an appendix for Chapter 7: "Empirical Investigation ..." "ASL-Speed 2020 Information Handout" are similar to appendix A.3 and "ASL-Speed 2020 Interview-Demographic-Paper-for-Participants" is similar to A.4.

B.1 Simple of the Selected Stories

- **Story1:** MANY PEOPLE THEY GO CAMPING FOREST VARIOUS STATES FIRST COLORADO SECOND WYOMING THIRD CALIFORNIA FOURTH WASHINGTON THEY SCARED WHY BLACK BEAR BROWN BEAR IF ATTACK DO THEY THINK SHOOT BUT SCIENTISTS UNIVERSITY ALASKA MAKE NEW CHEMICAL DEFENSE SPECIAL RED #PEPPER SPRAY AGAINST BEAR SHOO LAST YEAR RESEARCH EXPERIMENT SPRAY THERE RIFLE THERE COMPARE THERE STOP BEAR ATTACK #60 PERCENT SPRAY BETTER STOP #90 PERCENT ATTACK OTHER SCIENTISTS AFRICA MAKE SPRAY AGAINST INSECT READY WHEN NEXT YEAR

- **Story2:** RICE IT COST INCREASE NOW MANY COUNTRY WORRY NOT ENOUGH INDIA EGYPT LIMIT RICE SEND OTHER COUNTRY THERE RICE IMPORTANT FOOD FOR MANY

PEOPLE WORLD #3 MONTH PAST COST INCREASE DOUBLE WHY NONE RAIN RICE DISEASE SPREAD FINISH KILL MANY PLANTS ALSO SOME FARMERS CHANGE PLANT DIFFERENT PLANTS FRUIT VEGETABLE NOW MANY POOR PEOPLE #ASIA CANNOT BUY GOVERNMENT WORRY PEOPLE PROTEST PEOPLE THEY GATHER HIDE RICE CHINA PRESIDENT #XIANG ORDER ARMY SEARCH THEM AMERICA HERE COST INCREASE #8 PERCENT

- **Story3:** #ALBERT HE UNIVERSITY STUDENT PAST HE CHILD INTEREST MANY VARIOUS HE LIKE DANCING PIANO WATCH #DVD AND LOVE CHAT AND LEAVE WITH FRIENDS #ALBERT MANY BOYS SAME BUT HE KNOW WANT BECOME WRITER FOR BIG NEWSPAPER PAST HE CHILD INFORM MOTHER HE WANT BECOME WRITER MONTH FUTURE HE GRADUATE KNOW MUST PLAN HE READ ADVERTISEMENTS NEWSPAPER MAGAZINE COMPUTER TEACHER PARENTS GRANDFATHER ALL ADVISE HIM YESTERDAY HE NOTICE ADVERTISEMENT FIND PERFECT JOB

B.2 ASL-Speed 2020 Study Time Parameter Configuration

In this section of appendix B, I am presenting the different configuration used in the user study. This configuration could be used a reference for timing values in ASL for other researchers.

- **Sign Duration Parameter Configuration** We are assuming that Duration is a constant amount of time, we say X is the standard dictionary value for word.

Table B.1: Sign duration values

Parameter	**V. Slow**	**Slow**	**Normal**	**Fast**	**V. Fast**
Sign Duration	**4/X**	**2/X**	**X**	**X*2**	**X*4**
Transition	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pause Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Differential Rate	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pausing	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults

- **Transition Parameter Configuration** We are assuming that Transition is a constant unit of time, we say X is the standard SSS value of 0.25.

Table B.3: Transition values

Parameter	**V. Slow**	**Slow**	**Normal**	**Fast**	**V. Fast**
Sign Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Transition	**4*X**	**2*X**	**X**	**X/2**	**X/4**
Pause Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Differential Rate	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pausing	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults

- **Pause Duration Parameter Configuration** We are assuming that Pause Duration is a constant unit of time, we say X is the standard SSS value of 1.

Table B.5: Pause duration values

Parameter	**V. Slow**	**Slow**	**Normal**	**Fast**	**V. Fast**
Sign Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Transition	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pause Duration	**4*X**	**2*X**	**X**	**X/2**	**X/4**
Differential Rate	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pausing	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults

- **Pausing Frequency Parameter Configuration** We are building a rule to select each level of pausing frequency parameter.

Table B.7: Pausing frequency values

Parameter	**V. Slow**	**Slow**	**Normal**	**Fast**	**V. Fast**
Sign Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Transition	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pause Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Differential Rate	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pausing	**Pause on every word**	**Pause on verb phrase, noun phrase, clause, and sentences**	**Pause on verb phrase, clause, and sentences**	**Pause on every sentence**	**No Pauses**

- **Differential Rate Parameter Configuration** We are assuming that Differential Rate is a factor that multiplies by some time, we say X is the output of the machine learning model for Differential Rate modeling.

Table B.9: Differential rate values

Parameter	**V. Consistent**	**Consistent**	**Normal**	**Deviant**	**V. Deviant**
Sign Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Transition	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Pause Duration	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults
Differential Rate	$\mathbf{X^{0.25}}$	$\mathbf{X^{0.75}}$	**X**	$\mathbf{X^{1.5}}$	$\mathbf{X^{2}}$
Pausing	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults	SSS Defaults

B.3 ASL-Speed 2020 Advertisement

RIT

Advertisement for Research Study
Earn $40 for a one-hour appointment
Open to people who used ASL since childhood

When you were a child, did you use American Sign Language at home?

When you were a child (before age 7), did you use American Sign Language at school?

If you answer "yes" to one of these questions, you are invited to participate in a research study at Rochester Institute of Technology. Participants will view animations or videos of sign language for one hour on a computer, and they will be paid $40 for their participation.

The computer animations are of American Sign Language sentences, and participants will fill out a survey about whether the animations are correct and understandable.

To make an appointment, please contact:

Becca Dingman, Research Assistant
Linguistic and Assistive Technologies Laboratory
Rochester Institute of Technology
Email: bad6955@rit.edu

For general information about the project, please contact:

Dr. Matt Huenerfauth, Professor
Department of Information Sciences and Technologies
Rochester Institute of Technology (RIT)
Email: matt.huenerfauth@rit.edu

B.4 Interview Plan and Questions

Code: _____
ASL-Animation-vs-Human-2020-Interview

INTERVIEW CHECKLIST (Take notes during the interview.)

1. INFORMED CONSENT HANDOUT

2. FIRST TWO PAGES OF DEMOGRAPHIC PAPER

3. Run the study website **page1.html**
4. You begin the interview with some warm-up questions, about whether they had seen animations of ASL before, and what they had thought of them.

Move to next page (**#1 Sign Duration**)
5. Begin the interview by showing them the **Sign Duration** set of animations and explain to the participants that now he\she should focus **Sign Duration**. In **Sign Duration** videos you can explain to the participant that the signs are all the same, and the only change in these videos is **the duration or length** of the signs. Also, you should tell the participants that the signs in the example image **will be different** that the signs they will see in the animations on the next screen. (It is a different story.)

Let the participants watch the videos and answer the question on the two consecutive pages about **#1 Sign Duration**. Then stop by the discussion page, and discuss with the participants about the following: **(PLEASE TAKE NOTES HERE.)**

How important is this issue of **Sign Duration**?

Any positive comments about the sign durations? Are there some you liked? Why?

Any negative comments about the sign durations? Did you dislike some? Why?

Any suggestions about what we can do better?

6. The show them the **Transition** set of animations. Explain to the participants that now they should focus **the time that the hands move in-between signs**. Please point out to the participant that the signs **have the same duration/width** in each row, but the time in-between signs is different. This time in-between signs is illustrated by the gradient between the rectangles representing the signs. *Let the participants watch the videos and answer the question on the two consecutive pages.* Then stop by the discussion page, and ask the participants about: **(PLEASE TAKE NOTES HERE.)**

Is the time in-between the signs important for making the signing clear?

Any positive comments about the transition speeds? Are there some you liked? Why?

Any negative comments about the transition speeds? Did you dislike some? Why?

Any suggestions about what we can do better?

7. Then, show them the **Pausing** set of animations. Explain to the participants that now they should focus on some moments during signing when there is a **Pause** between signs, for example at the end of sentences or at other places. When showing them the drawing, you can point out that all the words, speed of words, and other details are identical – the only difference is that sometimes there are Pauses inserted during the story. Some have more Pauses, and some have fewer Pauses.

Let the participants watch the videos and answer the question on the two consecutive pages. NOTE: We anticipate that participants may be slower when they are watching the videos for this category (**Pausing**) and the next category (**Pause Duration)** of videos**.** Because pauses are less frequent, the participant may need to spend more time watching the animation before they form an opinion. So, you can let them know that it is OK to watch these videos carefully. Also, you should give the participant some advice before they see the video: You can explain to the participants that they may need to watch a longer portion of the video in order to see a **PAUSE.**

When done with the questions onscreen, ask the participants some questions:
(PLEASE TAKE NOTES HERE.)

Ask them to take a careful look at the videos. Is it easy to see where the sentences begin and end? Does it look like what a human would do? Are they natural?

Are pauses important? Do they make it easier to remember chunks of information?

Any positive comments about the pauses? Are there some videos you liked? Why?

Any negative comments about the pauses? Did you dislike some videos? Why?

Any suggestions about what we can do better?

8. Then, show them the **Pause Duration** set of animations. Explain to the participants that now they should focus on how long the pauses are between signs. In the drawing, you should point out that the signs, the speed of signs, and the location of Pauses is identical for all the rows in the drawing. Only the length of the pauses will change.

Let the participants watch the videos and answer the question on the two consecutive pages. NOTE: We anticipate that participants may be slower when they are watching the videos for this category (**Pause Duration**) and the prior category (**Pausing).** Because pauses are less frequent, the participant may need to spend more time watching the animation before they form an opinion. So, you can let them know that it is OK to watch these videos carefully. Also, you should give the participant some advice before they see the video: You can explain to the participants that they may need to watch a longer portion of the video in order to see a **PAUSE. And they can focus on how long the pause is. Do they like this?**

Then, ask the participants and take note about the following: (**PLEASE TAKE NOTES**)

Is the duration of pauses important? Do they make it easier to remember chunks of information? Do they make it look more natural? Help to show the structure?

Any positive comments about the pause duration? Are there some videos you liked? Why?

Any negative comments about the pause duration? Did you dislike some videos? Why?

Any suggestions about what we can do better?

9. Then, show them the **Differential Rate** set of animations. When showing the participant the drawing, explain to the participants that now they should focus on the overall **speed of signing**. In the picture, **everything gets faster or everything gets slower.** *Let the participants watch the videos and answer the question on the two consecutive pages.*

Then, ask the participants and take note about: (**PLEASE TAKE NOTES**)

Is the speed important? Are some words or phrases easy or difficult to see? Are some videos boring or slow? Are some too fast?

Any positive comments about the speed? Are there some videos you liked? Why?

Any negative comments about the speed? Did you dislike some videos? Why?

Any suggestions about what we can do better?

Ask if there are particular words they notice are hard to see or too slow?

Is it easy to tell where the sentences begin and end?

---

10. Rank the importance of the above five timing parameters (**A. Sign Duration, B. Transition, C. Pausing, D. Pauses duration, E. Differential Rate**) from very important to less important?

---

11. Broaden the discussion to focus on **animations of ASL in general**, not just this specific one shown in this study.

What types of speed would they like to see in animations like this in general. Are you worried about them being too fast or too slow? If other animations looked like this, would it be good?

Are there times when they would want more robotic signing or more natural signing?

Do they think pauses and speed are important?

If they try to sign these sentences themselves, how is it different?

What do you like about computer animations, in general?

What do you think about ASL animations?

What do you think is the most important characteristics for computer animation of ASL? Why?

(As needed, you can explain that our lab is NOT trying to use these to replace interpreters. The idea is that companies usually have websites in lots of languages, but never ASL because videos of real people are too hard to keep updated. We're trying to make it easier and cheaper for companies to put ASL on websites, but a Deaf person or an interpreter would still need to write the message that the animation shows. It would just be easier for the animation message to be updated, rather than videos of people. Our goal is to get websites of companies and governments around the world to have an ASL version, just like they have English and Spanish versions.)

---

12. Broaden the discussion, to focus on speed and pauses in **HUMAN** signing.

How does the participant use speed or pauses themselves?

What speed or pausing do they look for in the signing of other people.

If they see someone who is pausing in weird places or has unusual speed changes, do they seem like a "beginner" signer? Or a child? What would they thing if this stuff was wrong.

13. Finally, you can **wind-down the interview**. You can ask about any other feedback they have. You can ask them to mention something they likes and something we should update?

14. And then get them to sign a receipt when you pay them $40 at the end.

Appendix C

Appendix for Other Contributions

This is an appendix for Chapter 8: "Investigating Acceleration Curves in ASL. In this appendix, I am presenting my other contributions on a different project during my PhD study. These works are not explicitly mentioned in this dissertation.

Project 1 *"Gaze-guided Magnification for Individuals with Vision Impairments"* published in [cw]

Abstract: Video-based eye trackers increasingly have the potential to improve on-screen magnification for low-vision computer users. Yet, little is known about the viability of eye-tracking hardware for gaze-guided magnification. We employed a magnification prototype to assess eye tracking quality for low-vision users as they performed reading and search tasks. We show that a high degree of tracking loss prevents current video-based eye tracking from capturing gaze input for low-vision users. Our findings show current technologies were not made with low vision users in mind, and we offer suggestions to improve gaze-tracking for diverse eye input.

Project 2 *"Gaze Guidance for Captioned Videos for DHH Users"* published in [cv]

Abstract: Automatic Speech Recognition (ASR) technology can generate real-time captions during classroom lectures for Deaf or Hard-of-Hearing (DHH) individuals, but the resulting experience may not be fully accessible due to errors in captions (Berke et al. 2018) or the challenges faced when users must split their attention between the caption and other visual information, e.g. slides displayed (Kushalnagar et al. 2010). This study focuses on students viewing captioned videos of lectures containing an instructor and other visual content. We investigate whether the addition of gaze guidance (brief subtle blinking elements added to the video to draw someone's gaze) could be used to guide the visual attention of DHH individuals toward regions of the video where key information may be displayed. To avoid the time-consuming work of manually identifying specific times and locations in the video when we should guide the DHH users' gaze away from captions, we sought to automate this process by analyzing where hearing individuals (people who learned English as a second language) directed their gaze when they viewed these videos. The main contribution of this work is empirical: We have explored whether gaze guidance added to educational lecture videos lead to differences in DHH user's looking at the non-caption region of the video or in their comprehension of the content.

Project 3 *"Evaluating Sign Language Animation through Models of Eye Movements"* published in [al]

Abstract: We investigate whether machine learning algorithms can be trained on eye-tracking data from people who watch ASL animations, to predict whether the person watching the animation judges it to be of high-quality or easy to understand. As discussed in (Huenerfauth and Kacorri 2016), the advantage of this approach is that researchers do not need to design comprehension questions specifically tailored to the information content of the animations shown. Furthermore, by analyzing eye-movements rather than asking overt questions, researchers can avoid artificially drawing participants' attention to specific aspects of the animation, e.g. with questions about particular facial expressions, which could change how the participant views the animation.

Project 4 *"Design and Evaluation of a User-Interface for Authoring Sentences of American Sign Language Animation"* published in [am]

Abstract: We investigate the design of a user-interface for authoring sentences or multi-sentence messages in American Sign Language (ASL), using an animation platform at our lab that

includes a collection of pre-built ASL signs. In this formative study, participants expressed a preference for a "timeline" layout for arranging words to create a sentence, with a dual view of the word-level and the sub-word "pose" level.

Project 5 *"Modeling the Use of Space for Pointing in American Sign Language Animation"* published in [an]

Abstract: In ASL, signers associate items under discussion with locations around their body (McBurney 2002; Meier 1990), which the signer may point to later in the discourse to refer to these items again. For instance, if a signer were discussing a favorite book, she might mention the title of the book once, and then point at a location in space around her body. For the remainder of the conversation, she would not mention the title of the book again, but instead she would point to this location in space to refer to it. In this work, we model and predict the most natural locations for these spatial reference points (SRPs), based on recordings of human signers' movements. We evaluated ASL animations generated from the model in a user-based study.

Project 6 *"Evaluation of an English Word Look-Up Tool for Web-Browsing with Sign Language Video for Deaf Readers"* published in [ao]

Abstract: We have designed an interface to assist these users in reading English text on web pages; users can click on certain marked words to view an ASL sign video in a pop-up. A user study was conducted to evaluate this tool and compare it with web pages containing only text, as well as pages where users can click on words and see text-definitions using the Google Dictionary plug-in for browsers. The study assessed participants' subjective preference for these conditions and compared their performance in completing reading comprehension tasks with each of these tools. We found that participants preferred having support tools in their interface as opposed to none, but we did not measure a significant difference in their preferences between the two support tools provided.

Appendix D

Appendix for Annotation ELAN corpus

This is an appendix for the annotation guide used to build the ELAN version of the corpus. The annotation guide can be found on this link: ELAN-Annottation-Online.pdf.

Appendix E

Publications

- **Sedeeq Al-khazraji**, Becca Dingman, Sooyeon Lee, Matt Huenerfauth. 2021. At a Different Pace: Evaluating Whether Users Prefer Timing Parameters in American Sign Language Animations to Differ from Human Signers' Timing. The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'21). ACM, New York, NY, USA.

- Qiwen Zhao, Vaishnavi Mande, Paula Conn, **Sedeeq Al-khazraji**, Kristen Shinohara, Stephanie Ludi, Matt Huenerfauth. 2020. "Comparison of Methods for Teaching Accessibility in University Computing Courses". The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20). ACM, New York, NY, USA.

- **Sedeeq Al-khazraji**, Becca Dingman, Matt Huenerfauth. 2020. Empirical Investigation of Users' Preferred Timing Parameters for American Sign Language Animations. In Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI'20 Extended Abstracts). ACM, New York, NY, USA.

- Natalie Maus, Dalton Rutledge, **Sedeeq Al-khazraji**, Kristen Shinohara, Reynold Bailey, Cecilia Ovesdotter Alm. 2020. Gaze-guided Magnification for Individuals with Vision Impairments. In Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI'20 EA). ACM, New York, NY, USA.

- Jessica Li, Matt Luettgen, Matt Huenerfauth, **Sedeeq Al-khazraji**, Reynold Bailey, Cecilia Ovesdotter Alm. 2020. Gaze Guidance for Captioned Videos for DHH Users. Journal on Technology and Persons with Disabilities, Volume 8, California State University, Northridge.

- **Sedeeq Al-khazraji**. 2019. Building Predictive Models for Modeling Speed and Timing of American Sign Language to Generate Realistic Animations. The 21th International ACM SIGACCESS Conf. on Computers and Accessibility (ASSETS '19), Doctoral Consortium Presentation and Poster Session. Pittsburgh, PA, USA.

- Abhishek Kannekanti, **Sedeeq Al-khazraji**, and Matt Huenerfauth. 2019. Design and Evaluation of a User-Interface for Authoring Sentences of American Sign Language Animation. In: Antona M., Stephanidis C. (eds) Universal Access in Human-Computer Interaction. Theory, Methods and Tools. HCII 2019. Lecture Notes in Computer Science, Volume 11572. Springer, Cham. ***2019 UAHCI Best Paper award.***

- Abhishek Mhatre, **Sedeeq Al-khazraji**, Matt Huenerfauth. 2019. Evaluating Sign Language Animation through Models of Eye Movements. Journal on Technology and Persons with Disabilities, California State University, Northridge.

- **Sedeeq Al-khazraji**, Larwan Berke, Sushant Kafle, Peter Yeung and Matt Huenerfauth. 2018. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18). ACM, New York, NY, USA, 259-270. ***2018 SIGACCESS Best Paper award. (26% acceptance rate)***.

- **Sedeeq Al-khazraji**. 2018. Using Data-Driven Approach for Modeling Timing Parameters of American Sign Language. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). ACM, New York, NY, USA, 497-500.

- **Sedeeq Al-khazraji**, Sushant Kafle, and Matt Huenerfauth. 2018. Modeling and Predicting the Location of Pauses for the Generation of Animations of American Sign Language. In Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, The 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan.

- Jigar Gohel, **Sedeeq Al-khazraji**, Matt Huenerfauth. 2018. Modeling the Use of Space for Pointing in American Sign Language Animation. Journal on Technology and Persons with Disabilities, California State University, Northridge.

- Dhananjai Hariharan, **Sedeeq Al-khazraji**, Matt Huenerfauth. 2018. Evaluation of an English Word Look-Up Tool for Web-Browsing with Sign Language Video for Deaf Readers. Universal Access in Human-Computer Interaction. Methods, Technologies, and Users. UAHCI 2018. Lecture Notes in Computer Science, Volume 10907, pages 205-215. Springer, Cham.